



HAL
open science

Rôle des mécanismes d'autorégulation dans la soumission à l'autorité

Johan Lepage

► **To cite this version:**

Johan Lepage. Rôle des mécanismes d'autorégulation dans la soumission à l'autorité. Psychologie. Université Grenoble Alpes, 2017. Français. NNT : 2017GREAH027 . tel-01769505

HAL Id: tel-01769505

<https://theses.hal.science/tel-01769505>

Submitted on 18 Apr 2018

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Psychologie sociale expérimentale**

Arrêté ministériel : 25 mai 2016

Présentée par

Johan LEPAGE

Thèse dirigée par **Laurent BÈGUE, Professeur, Université Grenoble Alpes**, et
codirigée par **Martial MERMILLOD, Professeur, Université Grenoble Alpes**

préparée au sein du **Laboratoire InterUniversitaire de Psychologie**
dans **l'École Doctorale des Sciences de l'Homme, du Politique et du Territoire**

Rôle des mécanismes d'autorégulation dans la soumission à l'autorité

Thèse soutenue publiquement le **4 décembre 2017**,
devant le jury composé de :

Mme Moïra MIKOLAJCZAK

Professeure, Université catholique de Louvain, rapporteure

M. Jean ROSSIER

Professeur, Académie des sciences, rapporteur

Mme Aïna CHALABAEV

Maître de conférences, Université Grenoble Alpes, examinatrice

M. Martial FOUCAULT

Professeur, Sciences Po - CEVIPOF, président du jury

M. Laurent BÈGUE

Professeur, Université Grenoble Alpes, directeur

M. Martial MERMILLOD

Professeur, Université Grenoble Alpes, directeur



Johan LEPAGE

**Rôle des mécanismes
d'autorégulation dans
la soumission à l'autorité**

Thèse dirigée par **Laurent BÈGUE** et codirigée par **Martial
MERMILLOD**



« Tout comme chez les automates, il y a sûrement une modification de l'équilibre interne qui se réduit sans aucun doute à des changements dans les schémas du fonctionnement neural. Les inhibiteurs et les désinhibiteurs chimiques compromettent la possibilité d'action de certains mécanismes neuraux, mais dans quelle mesure exacte, nous sommes techniquement incapables de le déterminer au niveau chimio-neurologique. Nous avons cependant accès à une expression phénoménologique de ce délicat changement parce qu'il entraîne une modification d'attitude. »

Milgram, 1974, pp. 166-167



Tout mon amour à ma maman, qui me manque tellement...

1960 – 2017

À toi qui bat le pavé depuis le Printemps 2016

REMERCIEMENTS

Tout d'abord, un grand merci à mon directeur de thèse Laurent Bègue pour la constance de son engagement et de son soutien malgré l'adversité. Un grand merci également pour ta confiance et ta patience.

À Martial Mermillod, un remerciement tout spécial pour son soutien indéfectible, sa gentillesse, et sa disponibilité au cours de ces dernières années.

Je remercie également Moïra Mikolajczak et les membres de son équipe pour la qualité de leur accueil, Aurélia Bugajska et Laura Ferreri pour leur collaboration enthousiaste, Sylvain Harquel et Pierre Chausse pour leur aide précieuse.

Pour son amitié et son aide, je remercie mon camarade de thèse, Oulmann Zerhouni.

Un grand merci à celles et ceux qui ont contribué à éclairer cette période pas toujours facile par leur gentillesse, leur humour, leur générosité, leur intelligence : Richard Monvoisin, la bande du Cortecs, Brice Beffara, Amélie Bret, Nicolas Morgado, Delphine Papillon, les zygotos de la Grange.

Je tiens aussi à remercier celles et ceux qui me rappellent sans cesse que l'Utopie est belle : Maxime et Marion, Guillaume, Tiphaine, Aude, Chris, Maëva, Anna, et les ami.e.s pirates. Clin d'œil complice à Camille.

Enfin je remercie ma famille pour son amour et son soutien sans faille, en particulier ma mère, extraordinaire femme djok. Ton soleil brille dans le cœur de ton ptit gars...

Et pour tout, ma doudou.

RESUME

Dans les expériences de Milgram sur la soumission à l'autorité (Milgram, 1963, 1965, 1974), les participants se sont vus ordonner d'administrer une série de chocs électriques d'intensité croissante à un autre participant (en réalité compère de l'expérimentateur) au nom d'une étude sur les effets de la punition sur l'apprentissage. Les résultats montrent que 62.5% des participants ont été jusqu'à infliger plusieurs chocs potentiellement mortels (condition standard ; Milgram, 1974). Ces résultats ont suscité un fort intérêt et sont toujours largement cités pour expliquer certains comportements destructeurs comme les actes de torture et de barbarie. Mais les travaux de Milgram ont également provoqué une forte controverse éthique et toute possibilité de réplique a été rapidement proscrite. Dans ce contexte, peu d'études expérimentales ont été réalisées et la question des mécanismes responsables de l'obéissance destructrice (OD) demeure sans réponse. La recherche récente a pu relancer l'étude expérimentale de l'OD par l'usage d'environnements immersifs. Ainsi, une récente étude IRMf reposant sur l'utilisation d'une version virtuelle du paradigme de Milgram montre que l'observation de la douleur de la victime dans ce contexte provoque un état de détresse personnelle chez les participants (*i.e.*, réaction émotionnelle aversive centrée sur soi). Ce résultat suggère que l'OD pourrait être en partie la conséquence d'un défaut de régulation de la détresse provoquée par les mécanismes de résonance empathique. En nous appuyant sur la recherche récente en neurosciences sociales, nous avons fait l'hypothèse que la vulnérabilité au stress pourrait faciliter l'OD *via* l'exercice d'un contrôle inhibiteur sur la résonance empathique responsable d'une diminution de l'aversion pour l'atteinte à autrui. Nous avons réalisé six expériences visant (i) à examiner la relation entre tonus vagal (biomarqueur de la vulnérabilité au stress), autoritarisme de droite (prédicteur classique de l'OD) et OD, (ii) à manipuler expérimentalement la capacité des participants à exercer un contrôle inhibiteur durant la procédure d'obéissance, (iii) à explorer la relation entre ondes thêta (biomarqueur du contrôle inhibiteur) et OD, (iv) à examiner la relation entre OD et activité hémodynamique au niveau du cortex préfrontal ventromédian (incluant le cortex orbitofrontal) et du cortex préfrontal dorsolatéral, régions cérébrales fortement impliquées dans l'empathie et la cognition morale. L'obéissance a été mesurée à l'aide de l'« Immersive Video Milgram Obedience Experiment ». Dans ce qu'ils ont d'essentiel, nos résultats montrent : (i) que l'autoritarisme de droite et l'OD sont associés à un moindre tonus vagal, et que les participants obéissants ont exercé un effort cognitif couplé à une diminution du stress physiologique durant la procédure d'obéissance (études 2 et 3), (ii) que l'affaiblissement expérimental du contrôle inhibiteur *via* l'induction d'une « fatigue mentale » favorise la désobéissance et supprime l'influence de l'autoritarisme de droite sur l'OD (étude 4), (iii) que l'OD est associée à une augmentation de la puissance thêta (étude 5), (iv) qu'une augmentation de l'oxy-hémoglobine au niveau du cortex préfrontal ventromédian droit est associée à une moindre obéissance (étude 6). Dans leur ensemble, ces résultats supportent l'hypothèse voulant que la vulnérabilité au stress favorise l'exercice d'un contrôle inhibiteur sur la résonance empathique durant la procédure d'obéissance. Une moindre résonance empathique étant responsable d'une moindre aversion pour l'atteinte à autrui, les individus sont alors plus disposés à exécuter des ordres destructeurs.

Mots-clés : soumission à l'autorité, autoritarisme de droite, autorégulation, stress, contrôle inhibiteur, aversion pour l'atteinte à autrui

ABSTRACT

In the Milgram's obedience experiments (Milgram, 1963, 1965, 1974), naive participants were ordered to administer increasingly severe electric shocks on a “learner” (a confederate) after being told that they were participating in an experiment on the effects of punishment on learning. Results revealed that 62.5% of the participants were willing to administer allegedly lethal electric shocks when ordered to do so (standard condition; Milgram, 1974). The Milgram's findings are still often cited when explaining destructive behaviors such as torture. The Milgram's obedience studies have also been a target of ethical criticism and replication has been discouraged. In such a context, a very few experimental studies has been conducted since the Milgram's experiments and the mechanisms responsible for destructive obedience remain unknown. Recent research reopens the door to direct empirical study of destructive obedience through the employment of immersive environments. A recent fMRI study showed that pain-related affective sharing in a virtual version of the Milgram paradigm elicited an aversive, self-oriented state of personal distress. This result suggests that low self-regulatory control of the shared affect evoked by the victim's pain could be responsible for destructive obedience. Based on recent social neuroscience research, we hypothesized that stress vulnerability may facilitate destructive obedience through a mechanism of inhibitory control over empathic resonance responsible for decreased harm aversion. We conducted six studies aiming (i) to explore the relation between cardiac vagal tone (a biomarker of stress vulnerability), right-wing authoritarianism (RWA, a classic predictor of destructive obedience) and destructive obedience, (ii) to induce a self-regulatory fatigue in order to manipulate the participants' abilities for inhibitory control during the obedience procedure, (iii) to explore the relation between theta oscillations (a biomarker of inhibitory control) and destructive obedience, (iv) to examine the relation between destructive obedience and hemodynamic response in the ventromedial prefrontal cortex (including the orbitofrontal cortex) and the dorsolateral prefrontal cortex, two brain areas highly involved in empathy and moral cognition. Obedience was measured using the “Immersive Video Milgram Obedience Experiment”. All in all, our results showed: (i) that lower vagal tone was related to higher RWA and destructive obedience, and that obedient participants exerted a cognitive effort associated with decreased physiological arousal (studies 2 and 3), (ii) that self-regulatory fatigue reduced destructive obedience and suppressed the influence of RWA, (iii) that destructive obedience was related to increased theta power (study 4), (iv) that increased oxygenated-hemoglobin in the right ventromedial prefrontal cortex was related to disobedience. On the whole, these results support the hypothesis that stress vulnerability motivates the use of inhibitory control over empathic resonance during the obedience procedure. Because decreased empathic resonance is responsible for decreased harm aversion, individuals are more prone to obey destructive orders.

Keywords: obedience to authority, right-wing authoritarianism, self-regulation, stress, inhibitory control, harm aversion

Le rejet progressif du dualisme cartésien dans le domaine des sciences du vivant a permis la constitution de nouveaux paradigmes liant corps et esprit (e.g., Damasio, 1994). L'émergence et le développement d'approches comme la psychophysiologie sociale, les neurosciences sociales, la psychoneuroendocrinologie ou encore la psychoneuroimmunologie attestent de la pertinence d'une approche « multiniveaux » pour la compréhension du comportement humain. Du point de vue épistémologique, ces approches reposent sur la « *théorie de l'identité* » (Armstrong, 1968 ; Feigl, 1958 ; Place, 1956 ; Smart, 1959) proposant que la propriété d'un état mental (e.g., ressentir de la douleur) est identique avec une certaine propriété neurale. Il n'est pas question ici d'une simple corrélation entre état mental et état neural : « dire qu'ils sont *corrélés*, c'est dire qu'il y a quelque chose 'de plus' » (Smart, 1959). Par conséquent, la « *théorie de l'identité* » affirme que les propriétés mentales sont *identiques* avec des propriétés physiques : « vous ne pouvez pas corréler quelque chose à lui-même » (Smart, 1959). L'état mental et l'état neural sont ainsi strictement identiques : mentionner une propriété mentale revient à mentionner une propriété physique. Les structures biologiques et les processus physiologiques « incorporant » tous les aspects du fonctionnement humain (émotions, pensées, actions), les chercheurs peuvent porter leur attention sur ces structures et processus afin d'avancer dans leur compréhension du comportement humain (Cacioppo & Tassinary, 1990). Les neurosciences sociales constituent une approche intégrative et multidisciplinaire dans l'étude des mécanismes psychologiques. Cette démarche est une « extension naturelle » de la cognition sociale (Amodio & Ratner, 2013) dans la mesure où elle pose des questions similaires mais avec une approche étendue intégrant les composantes biologiques du comportement et les techniques de mesure physiologique comme l'imagerie cérébrale, l'électroencéphalographie, ou encore

l'électrocardiographie. Pour un chercheur en psychologie sociale intéressé par les bénéfices potentiels d'une approche neuroscientifique, la question de départ est de savoir si son intérêt porte sur des processus psychologiques qui réfèrent à des phénomènes ou à des mécanismes. Les modèles et outils développés par les neurosciences peuvent être d'une grande utilité dans le cas où les processus psychologiques d'intérêt sont réductibles à des mécanismes physiologiques « basiques » (e.g., Amodio & Ratner, 2013 ; Cacioppo & Tassinary, 1990 ; Dovidio, Pearson, & Orr, 2008). En ce qui concerne les techniques de mesure physiologique, celles-ci permettent des mesures objectives, continues et en temps réel de processus physiologiques largement automatiques et/ou incontrôlables associés à des construits motivationnels, affectifs et cognitifs – des construits difficiles à quantifier sans biais ou artefacts en l'absence de telles mesures et impossibles à quantifier autrement qu'*in vivo* (e.g., Blascovich, 2014 ; Blascovich & Mendes, 2010). Ainsi, pour des raisons à la fois méthodologiques, théoriques et épistémologiques, ce travail doctoral en psychologie sociale expérimentale reposera en grande partie sur une approche guidée aux niveaux théorique et méthodologique par les neurosciences sociales.

TABLE DES MATIERES

REMERCIEMENTS	11
RESUME	13
ABSTRACT	15
AVANT-PROPOS	17
INTRODUCTION	25
PARTIE 1. CADRE THEORIQUE	35
CHAPITRE 1. SOUMISSION À L'AUTORITE	37
1.1.1. Paradigme d'obéissance de Milgram	37
<i>Recrutement des participants, local et personnel</i>	37
<i>Méthode expérimentale</i>	37
<i>Tâche de mémorisation et test de rappel</i>	38
<i>Générateur de chocs électriques</i>	39
<i>Injonctions de l'expérimentateur</i>	40
<i>Réactions de l'élève</i>	40
<i>Mesure de l'obéissance</i>	41
<i>Débriefing</i>	41
<i>Prédictions comportementales</i>	41
1.1.2. Résultats (expérimentations 1-4)	44
<i>Réactions des participants</i>	47
<i>Aspects sociodémographiques et interculturels</i>	47
1.1.3. Déterminants situationnels et dispositionnels de l'OD	49
1.1.3.1. Déterminants situationnels de l'OD	49
<i>Méta-analyse de Haslam, Loughnan et Perry (2014)</i>	52
<i>Méta-analyse de Packer (2008)</i>	55
1.1.3.2. Déterminants dispositionnels de l'OD	56
<i>Problèmes méthodologiques</i>	56
<i>Problèmes psychométriques</i>	56
1.1.3.2.1. Influence de la personnalité et des attitudes sur l'OD	59
1.1.3.2.2. Autoritarisme	61
<i>Premières propositions théoriques</i>	61
<i>Altemeyer et l'échelle d'autoritarisme de droite</i>	63
<i>Recherche récente sur les origines du RWA</i>	64
	19

<i>Orientation à la dominance sociale</i>	65
<i>Nouvelles perspectives sur l'autoritarisme</i>	66
<i>Modèle duel motivationnel</i>	67
<i>Modèle de la sécurité collective</i>	70
<i>Recherche sur l'autoritarisme en contexte français</i>	71
1.1.3.2.3. Influence des compétences socio-émotionnelles sur l'OD	72
1.1.4. Modèles théoriques de l'OD	75
<i>Hypothèse de l'état agentique</i>	75
<i>Désengagement moral</i>	77
<i>L'OD, un comportement « négocié » ?</i>	80
<i>Hypothèse de l'identification sociale</i>	81
1.1.5. Approches discontinuiste et continuiste de la moralité humaine	83
<i>Approche discontinuiste</i>	83
<i>Approche continuiste</i>	84
CHAPITRE 2. APPORTS DES NEUROSCIENCES SOCIALES	87
1.2.1. Neurosciences de la morale	87
1.2.1.1. Approche dualiste de la morale	87
1.2.1.2. Substrats neurobiologiques	87
<i>Amygdale cérébrale</i>	90
<i>Cortex préfrontal ventromédian et cortex orbitofrontal</i>	90
<i>Cortex préfrontal dorsolatéral</i>	91
<i>Cortex cingulaire antérieur</i>	91
1.2.2. Neurosciences sociales de l'empathie	92
<i>Résonance neurale et affective face à la douleur d'autrui</i>	92
<i>Sympathie vs détresse personnelle</i>	95
<i>Modèle de Decety et Jackson (2004)</i>	96
<i>Contexte et similarité</i>	98
1.2.3. Première approche de l'OD par les neurosciences sociales	98
1.2.4. Mécanismes d'autorégulation	102
<i>Réponses émotionnelles primaires vs secondaires</i>	102
<i>Notion de contrôle volontaire</i>	104
<i>Stratégies d'autorégulation basées sur l'évitement</i>	106
<i>Notion de réévaluation cognitive</i>	108

1.2.5. Théorie polyvagale	110
1.2.5.1. Stress et homéostasie	110
1.2.5.2. Phylogénèse du SNA	111
<i>Principe jacksonien de dissolution</i>	113
<i>Neuroception du danger</i>	113
1.2.5.3. Mesure du tonus vagal	114
<i>Domaine temporel</i>	115
<i>Domaine fréquentiel</i>	115
1.2.5.4. Évidence empirique	116
1.2.6. Modèle d'intégration neuroviscérale	117
1.2.7. Neurosciences politiques	121
<i>Flexibilité cognitive</i>	122
<i>Traitement de la menace</i>	123
CHAPITRE 3. ENJEUX ETHIQUES ET METHODOLOGIQUES DANS LA RECHERCHE SUR L'OD	125
1.3.1. Coût méthodologique du recours à la duperie	125
<i>Suspicion dans les expériences de Milgram</i>	126
<i>Recherche sur les effets de la suspicion</i>	127
1.3.2. Jeu de rôle actif et OD	129
1.3.3. Technologie d'environnement immersif	130
« <i>Virtual Milgram paradigm</i> »	132
« <i>Immersive Video Milgram Obedience Experiment</i> »	136
PARTIE 2. PROBLEMATIQUE GENERALE	141
PARTIE 3. AUTORITARISME, UTILITARISME MORAL ET TONUS VAGAL	149
3.1. Autoritarisme et Utilitarisme moral	151
<i>Objectifs et présentation de l'étude</i>	151
Manuscrit 1	153
3.2. Autoritarisme et Tonus vagal	183
<i>Objectifs et présentation de l'étude</i>	183
Manuscrit 2	184
PARTIE 4. OBEISSANCE DESTRUCTRICE, AUTOREGULATION ET « FATIGUE MENTALE »	215
4.1. Obéissance destructrice et Tonus vagal	217

<i>Objectifs et présentation de l'étude</i>	217
Manuscrit 3	218
4.2. Obéissance destructrice et « fatigue mentale »	262
<i>Objectifs et présentation de l'étude</i>	262
Manuscrit 4	264
PARTIE 5. SUBSTRATS NEURONAUX DE L'OBEISSANCE DESTRUCTRICE	309
5.1. Étude EEG : Obéissance destructrice et contrôle inhibiteur	311
<i>Objectifs et présentation de l'étude</i>	311
Manuscrit 5	312
5.2. Étude fNIRS : Obéissance destructrice et intégration émotionnelle	346
<i>Objectifs et présentation de l'étude</i>	346
Manuscrit 6	348
PARTIE 6. DISCUSSION GENERALE ET CONCLUSION	385
DISCUSSION GENERALE	387
6.1. Rappel du contexte général	387
6.2. Questionnements directeurs	388
6.3. Principaux résultats	391
<i>Autoritarisme, utilitarisme moral et tonus vagal</i>	391
<i>Flexibilité physiologique et OD</i>	391
<i>Fatigue mentale et OD</i>	392
<i>Activité électroencéphalographique et OD</i>	392
<i>Taux d'oxy-hémoglobine au niveau du CPF et OD</i>	393
6.4. Implications théoriques	394
<i>Obstacles épistémologiques</i>	396
<i>Déterminisme biologique et déterminisme social</i>	396
6.5. Limites et perspectives	399
<i>Rôle causal du tonus vagal</i>	399
<i>Mimétisme neuromusculaire et OD</i>	401
<i>Sérotonine et OD</i>	404
CONCLUSION	406
BIBLIOGRAPHIE	409
ANNEXES	451
<i>Annexe 1 – CERNI – 4/07/2013</i>	451

<i>Annexe 2 – CPP – 13/02/2015</i>	452
<i>Annexe 3 – Notice d’informations études 3-6</i>	454
<i>Annexe 4 – Dernière page de la Notice d’informations étude 6</i>	459
<i>Annexe 5 – Procédure de recrutement études 3-6</i>	460
<i>Annexe 6 – Formulaire d’inclusion en ligne études 2-6</i>	461
<i>Annexe 7 – Formulaire d’inclusion études 5 et 6</i>	464
<i>Annexe 8 – Exemple de Flyer utilisé pour le recrutement</i>	465
<i>Annexe 9 - Page de Blog pour étude 4</i>	466
LISTE DES ABREVIATIONS	467
LISTE DES TABLEAUX	468
LISTE DES FIGURES	469

INTRODUCTION

Héritages évolutifs

L'étude des primates sociaux montre que tous les primates appartenant à la classe des hominidés (humain, chimpanzé commun, bonobo, gorille, orang-outan) possèdent un système social hiérarchique (e.g., Bercovitch, 1991 ; Leonard, 1979 ; Mazur, 1985 ; Sapolsky, 1993, 1995). Cette organisation hiérarchique présenterait un certain nombre d'avantages adaptatifs : (i) facilitation de la répartition des ressources et réduction des conflits au sein du groupe, (ii) optimisation des chances de survie pour le groupe (les groupes hiérarchisés développent des stratégies d'attaque et de défense plus efficaces que les groupes non hiérarchisés ; e.g., de Waal, 2003 ; Sidanius & Pratto, 1999). Les systèmes hiérarchiques des primates non humains se caractérisent par une structure trimorphique semblable à celle des humains avec une détermination du statut social sur la base de l'âge, du sexe et de critères arbitraires comme la position de la parenté au sein du groupe (e.g., Kawanaka, 1982, 1989 ; Nadler, 1988 ; Strier, 1994). Il existe d'autres manifestations chez les primates non humains de l'arbitraire hiérarchique, comme l'ajustement de l'agressivité de la progéniture en fonction du changement de rang social de la mère (Johnson, 1987) ou la formation de coalitions entre mâles de haut statut pour assurer leur domination (Harcourt, 1988 ; Leigh & Shea, 1995). Ainsi, l'organisation hiérarchique des sociétés humaines serait la conséquence de l'histoire évolutive de l'humain.

Du fait de cette permanence de l'organisation hiérarchique chez les hominidés, il est raisonnable de considérer l'*obéissance* comme une composante importante de la socialité qui s'est profondément ancrée dans la psychologie et la biologie des primates sociaux, humain inclus. Plusieurs auteurs ont ainsi proposé que l'évolution ait doté l'humain d'une prédisposition à se soumettre à l'autorité et plus largement à répondre positivement aux

symboles de l'autorité (e.g., Bouchard, 2009). Pour Somit et Peterson (1997), la pression évolutive aurait doté l'humain d'un « biais génétique » en faveur des formes d'organisation caractérisées par la hiérarchie, la dominance et l'obéissance. Cet héritage évolutif expliquerait le fait que tous les systèmes politiques (démocratiques et autoritaires) présentent une structure hiérarchique, de même que la prévalence des systèmes autoritaires et la fragilité des systèmes démocratiques.

Il semble également que l'évolution ait doté l'humain d'une prédisposition à éprouver une forte *aversion pour l'atteinte à autrui* (e.g., Blair, 1995, 2007 ; Crockett, Clark, Hauser, & Robbins, 2010 ; Decety & Svetlova, 2012 ; de Waal, 2003). Selon le modèle « perception-action » (Preston & de Waal, 2002), la perception d'une émotion active automatiquement les mécanismes neuronaux responsables de la génération de cette émotion. Ce système conduirait l'observateur à entrer en « résonance » avec l'état émotionnel d'autrui, *via* l'activation des représentations motrices et des réponses autonomes et somatiques associées. De nombreuses études en imagerie par résonance magnétique fonctionnelle (IRMf) ont confirmé l'existence d'un tel mécanisme de « résonance empathique » (e.g., Decety & Lamm, 2009), souvent considéré comme une base biologique au raisonnement moral orienté vers le bien-être d'autrui et au comportement prosocial (*i.e.*, actions destinées à venir en aide ou à apporter un bénéfice à autrui ; Decety & Svetlova, 2012). L'automatisme des mécanismes de résonance ainsi que la présence de cette capacité chez tous les primates sociaux (e.g., Preston & de Waal, 2002) autorisent à considérer l'empathie comme une composante centrale de la socialité qui, à l'instar de l'obéissance, s'est profondément ancrée dans la psychologie et la biologie des primates sociaux.

Problématique de l'obéissance destructrice

La capacité à développer des relations bénéfiques dans un contexte hiérarchique représente un avantage adaptatif (e.g., de Waal, 1982 ; Fiske, 1991). L'obéissance réduirait notamment la nécessité d'un recours à la force physique qui exposerait dominants et dominés à des risques d'atteintes corporelles. Mais si le comportement d'obéissance facilite les interactions quotidiennes au sein de groupes hiérarchiques, l'Histoire montre qu'il peut également avoir pour conséquence une atteinte sévère à autrui, comme dans le cas des actes de torture et de barbarie, des crimes de guerre et des génocides (e.g., Kelman, 2005). On parle alors d'obéissance destructrice (OD), définie comme *l'obéissance à des ordres menant à porter atteinte à la dignité et l'intégrité physique et/ou morale d'autrui* (Kelman & Hamilton, 1989). Comme le rapportent de nombreuses études socio-historiques, ces « crimes d'obéissance » sont universels et se retrouvent tant à l'époque ancienne que moderne et contemporaine : Guerre du Péloponnèse, Inquisition médiévale, crimes nazis, massacre de My Lai, Abou Ghraïb, etc. (e.g., Atran, 2003 ; Charlesworth, 2003 ; Kelman & Hamilton, 1989 ; Payen, 2012).

La situation d'OD peut être définie comme une situation de *dilemme social* mettant en tension l'évitement de l'atteinte à autrui et la nécessité perçue de se soumettre à l'autorité. Par conséquent, d'aucun pourrait faire l'hypothèse que seuls les individus présentant un défaut majeur et chronique de résonance empathique adopteraient un comportement d'OD. Ce trait étant définitoire de la psychopathie (Blair, 2007), seule une minorité d'individus exécuteraient alors des ordres destructeurs (la prévalence de la psychopathie dans la population générale est de 1% ; Hare & Neumann, 2008). Les motivations idéologiques sont également souvent invoquées pour expliquer les crimes d'obéissance. Selon Jost et Amodio (2012), l'idéologie est une « puissante force motivationnelle » pouvant conduire les êtres humains à commettre

des actes destructeurs au nom de systèmes de croyances abstraites. Ces deux hypothèses (prédisposition psychologique à l'atteinte à autrui et fanatisme idéologique) ont notamment été avancées pour expliquer les crimes nazis. Prenant le contre-pied de ces approches, la philosophe Hannah Arendt (1963) a proposé comme facteur causal l'obéissance « ordinaire ». Arendt a couvert le procès du criminel de guerre nazi Adolf Eichmann (ouvert en avril 1961) et a publié en 1963 un essai intitulé *Eichmann à Jérusalem. Rapport sur la banalité du mal*. La thèse défendue par Arendt a suscité une intense controverse et influence encore fortement la compréhension de l'OD. Selon l'auteure, Eichmann était non pas un idéologue motivé par une volonté génocidaire, ou encore un psychotique ayant fait « du crime un principe », mais un bureaucrate banal et consciencieux agissant en tant que partie d'une mécanique étatique et ce dans une simple dynamique d'obéissance. On peut lire ainsi : « Il eût été réconfortant de croire qu'Eichmann était un monstre (...) L'ennui avec Eichmann, c'est précisément qu'il y en avait beaucoup qui lui ressemblaient et qui n'étaient ni pervers ni sadiques, qui étaient, et sont encore, effroyablement normaux. Du point de vue de nos institutions et de notre éthique, cette normalité est beaucoup plus terrifiante que toutes les atrocités réunies, car elle suppose que ce nouveau type de criminel, tout *hostis humani generis* qu'il soit, commet des crimes dans des circonstances telles qu'il lui est impossible de savoir ou de sentir qu'il a fait le mal. » (Arendt, 1963)

C'est dans ce contexte historique et intellectuel que Stanley Milgram (1963, 1965, 1974) a initié la recherche expérimentale sur l'OD. Les résultats des expériences réalisées par Milgram démontrent que des personnes volontaires à une expérience présentée comme une recherche sur la mémoire et l'apprentissage sont en grande majorité susceptibles d'accomplir des actes aussi extrêmes qu'infliger à autrui une série de chocs électriques potentiellement mortels sur ordre d'une autorité (62.5% dans la condition standard ; Milgram, 1974). Ces résultats, *a priori* cohérents avec l'hypothèse d'Arendt et répliqués dans plus d'une dizaine de

pays auprès de plus de 3000 personnes, ont provoqué un émoi important au sein de la communauté académique et extra-académique. Aujourd'hui, les travaux de Milgram font encore référence pour l'explication des actes de torture et de barbarie, des crimes de guerre et des génocides (e.g. Kelman, 2005 ; Lankford, 2009 ; Waller, 2002). Commentant l'impact sans précédent des résultats de Milgram, Ross et Nisbett (1991) ont désigné les expériences de Milgram comme faisant dorénavant partie de l'héritage intellectuel partagé des sociétés occidentales.

Toutefois, la question des *mécanismes responsables* de l'OD demeure sans réponse (e.g., Cheetham, Pedroni, Antley, Slater, & Jäncke, 2009). La forte controverse éthique provoquée par les travaux de Milgram et l'évolution des critères éthiques encadrant la recherche ont rapidement proscrit toute possibilité de réplique des expériences de Milgram. Dans ce contexte, la question des causes sous-jacentes à l'OD fait encore débat, et les données existantes sont régulièrement réinterprétées (e.g., Blass, 1991 ; Burger, Girgis, & Manning, 2011 ; Haslam, Loughnan, & Perry, 2014 ; Reicher & Haslam, 2011 ; Packer, 2008).

Rationalité et situation

Les tentatives pour comprendre les causes de l'OD présentent deux caractéristiques : (i) elles s'ancrent majoritairement dans une approche « situationniste » (*i.e.*, le comportement serait entièrement déterminé par la situation), (ii) elles se focalisent sur le rôle de la cognition et de la rationalité au détriment des émotions comme guide du comportement (e.g., Bandura, 1986 ; Milgram, 1974). Cette modélisation du fonctionnement humain a eu des répercussions sur la recherche : (i) les travaux de Milgram et ses continuateurs ont prioritairement porté sur l'influence de facteurs situationnels sur l'OD, (ii) les études reposent majoritairement sur l'usage d'outils de mesure explicite (*i.e.*, questionnaires et échelles).

Or, les expériences de Milgram mettent en lumière l'existence de *fortes différences interindividuelles* dans les réactions à des ordres destructeurs (e.g., Bègue et al., 2015 ; Berkowitz, 1999 ; Blass, 1991). Si la question des mécanismes responsables de ces différences demeure sans réponse, la littérature suggère que l'*autoritarisme de droite* serait un prédicteur robuste de l'OD (e.g., Altemeyer, 1981 ; Blass, 1991 ; Elms & Milgram, 1966). De plus, la recherche récente en neurosciences sociales suggère que la décision morale dans des situations impliquant une atteinte à autrui serait prioritairement déterminée par des *processus émotionnels automatiques* associés à la résonance empathique (e.g., Blair, 2007 ; Decety & Svetlova, 2012 ; Crockett et al., 2010 ; Funk & Gazzaniga, 2009; Moll & de Oliveira-Souza, 2007). Les modèles les plus récents de l'empathie mettent également en avant l'importance des *mécanismes d'autorégulation* dans l'initiation d'un comportement prosocial face à autrui en détresse : un état d'hyper-activation physiologique induit des réactions émotionnelles aversives centrées sur soi et la motivation à réduire non pas la détresse d'autrui mais son propre inconfort (e.g., Decety & Lamm, 2009 ; Eisenberg & Eggum, 2009).

Rôle des mécanismes d'autorégulation

Le développement de la technologie d'environnement immersif a permis la réouverture de l'étude expérimentale de l'OD au début des années 2000. Les résultats d'une récente étude IRMf réalisée à l'aide d'une version virtuelle du paradigme de Milgram suggèrent que le dilemme social induit par le paradigme d'obéissance pourrait favoriser l'expérience de la détresse personnelle (Cheetham et al., 2009). Cette étude offre un « point de départ » pour une investigation plus précise du rôle des mécanismes d'autorégulation dans l'OD. L'autorégulation dépend de manière critique de la « flexibilité physiologique » des individus (*i.e.*, capacité à ajuster son niveau d'excitation physiologique à un instant précis ; Gross, 1998). La recherche empirique montre que le tonus vagal reflète la flexibilité

physiologique des individus et constitue un robuste biomarqueur de la vulnérabilité au stress. Notamment, des études ont mis en évidence qu'un moindre tonus vagal favorise la détresse personnelle et l'adoption de stratégies d'autorégulation basées sur l'évitement, comme le contrôle inhibiteur des réponses internes (e.g., Aldao, Dixon-Gordon, & De Los Reyes, 2015 ; Fabes, Eisenberg, & Eisenbud, 1993).

Questions directrices

La question centrale de cette thèse porte sur le rôle des mécanismes d'autorégulation dans une situation d'OD. Dans le cadre d'une approche biopsychosociale, nous nous sommes demandé si les variations dans la vulnérabilité au stress président aux différences interindividuelles dans l'OD. Les interrogations suivantes ont guidé notre réflexion : (i) Les différences interindividuelles dans la flexibilité physiologique sous-tendent-elles les différences interindividuelles dans l'OD ? (ii) L'influence de l'autoritarisme de droite sur l'OD pourrait-elle s'expliquer par l'existence d'une relation entre autoritarisme de droite et flexibilité physiologique ? (iii) L'exercice d'un contrôle inhibiteur sur les réponses émotionnelles associées à la résonance empathique pourrait-il expliquer l'influence de la flexibilité physiologique sur l'OD ? (iv) Ce contrôle inhibiteur pourrait-il diminuer l'aversion pour l'atteinte à autrui et ainsi favoriser l'OD ? (v) Est-il possible d'influencer l'OD en manipulant la capacité des individus au contrôle inhibiteur ?

Ce travail doctoral a été divisé en six parties. La **partie 1** est consacrée au cadre théorique de notre recherche. Dans le **chapitre 1**, nous passerons en revue les principaux travaux réalisés sur l'OD. Nous présenterons dans un premier temps le paradigme d'obéissance élaboré par Milgram et les résultats des principales études de l'auteur. Nous aborderons ensuite les études portant sur le rôle de certains facteurs situationnels et dispositionnels. Cette partie sera l'occasion de mettre en évidence la pauvreté de la recherche

sur les déterminants de l'OD, mais également l'influence importante que semble exercer l'autoritarisme de droite sur l'OD. Nous accorderons une place importante à la recherche en psychologie sociale sur l'autoritarisme afin d'y trouver des éléments permettant d'expliquer cette influence. Nous terminerons ce chapitre par la présentation des principaux modèles théoriques de l'OD. Nous mettrons en exergue la sous-estimation du rôle des émotions dans la détermination du comportement humain et replacerons cette approche dans le cadre des divisions théoriques caractérisant l'étude de la moralité humaine. Le **chapitre 2** sera dédié aux apports des neurosciences sociales à la compréhension des réactions affectives-motivationnelles à la douleur d'autrui. Dans un premier temps, nous mettrons en évidence le rôle des processus émotionnels automatiques dans le jugement moral avant de présenter de manière plus extensive les travaux portant sur le rôle respectif de la résonance empathique et des mécanismes d'autorégulation dans les réactions à la douleur d'autrui. Nous terminerons avec la présentation de l'étude IRMf de Cheetham et al. (2009). Dans un deuxième temps, nous nous focaliserons sur la question de l'autorégulation. La notion d'autorégulation sera clarifiée à partir d'une distinction entre « réponse émotionnelle primaire » et « réponse émotionnelle secondaire », avant une présentation des travaux relatifs à l'influence des caractéristiques des individus et du contexte sur les réponses à la douleur d'autrui. Dans un troisième temps, nous présenterons les travaux théoriques et empiriques portant sur le rôle du système nerveux autonome dans la détermination des réponses aux défis posés par notre environnement. Nous accorderons une place importante à la théorie polyvagale développée par Porges (1995, 1998, 2001, 2003, 2007), théorie conférant un rôle central à la régulation de la fonction cardiaque par le nerf vague dans la sélection des réponses aux expériences stressantes. Nous terminerons ce chapitre par la présentation de travaux réalisés dans le champ émergent des neurosciences politiques suggérant une relation étroite entre autoritarisme et autorégulation. Le **chapitre 3** sera consacré aux enjeux méthodologiques liés

à l'étude expérimentale de l'OD. Nous aborderons dans un premier temps les problèmes éthiques mais aussi méthodologiques associés à l'usage de la duperie. Nous mettrons en évidence le fait que la duperie en contexte expérimental pourrait favoriser la suspicion des participants et de ce fait diminuer le contrôle expérimental. Dans un deuxième temps, nous présenterons la méthode du jeu de rôle actif comme alternative à la duperie. Nous terminerons par la présentation des avantages méthodologiques de la technologie d'environnement virtuel immersif et des études ayant eu recours aux environnements immersifs pour l'étude de l'OD.

La **partie 2** expose la problématique générale de ce travail. Les **parties 3-5** réunissent les six études que nous avons réalisées. La **partie 3** comprend deux études réalisées en amont de nos expériences sur la soumission à l'autorité portant sur la relation que l'autoritarisme pourrait entretenir avec l'utilitarisme moral et la flexibilité physiologique. Plus précisément, l'étude 1 ($n = 500$) interroge la relation entre autoritarisme et réponse à différentes versions du « dilemme du tramway », et l'étude 2 ($n = 119$) investigate la relation entre autoritarisme et tonus vagal.

La **partie 4** comprend deux études sur la soumission à l'autorité. L'étude 3 ($n = 85$) porte sur la relation entre flexibilité physiologique et OD. Afin de tester l'hypothèse d'une influence de la flexibilité physiologique sur l'OD, nous avons cherché à savoir si le tonus vagal et la réactivité vagale des participants prédisent leur comportement d'obéissance. Nous avons également tenté de tester l'hypothèse d'un effort au contrôle inhibiteur chez les participants obéissants (*i.e.*, augmentation du tonus vagal et diminution de la fréquence cardiaque). Des hypothèses de médiation ont également été testées (e.g., médiation de l'influence du tonus vagal sur l'OD par le contrôle inhibiteur). L'étude 4 ($n = 99$) avait pour objectif de manipuler la capacité des participants au contrôle inhibiteur en amont de la

procédure d'obéissance. Nous nous attendions à ce que l'induction d'une « fatigue mentale » (e.g., Francis & Inzlicht, 2016) favorise la désobéissance.

La **partie 5** rassemble deux études portant sur les substrats neuronaux de l'OD. Pour l'étude 5 ($n = 50$), nous avons eu recours à l'électroencéphalographie afin de pouvoir mettre en relation les variations de potentiels électriques engendrées par l'activité cérébrale des participants durant la procédure d'obéissance avec l'OD. Nous nous sommes particulièrement intéressés aux ondes thêta du fait de leur implication dans le contrôle inhibiteur (e.g., Ertl, Hildebrandt, Ourina, Leicht, & Mulert, 2013). Pour l'étude 6 ($n = 46$), nous avons eu recours à la spectroscopie proche infrarouge afin de mettre en relation l'OD avec l'activité hémodynamique au niveau du cortex préfrontal. Nous nous sommes particulièrement intéressés à l'activité hémodynamique au niveau du cortex préfrontal ventromédian et du cortex orbitofrontal, la recherche ayant démontré qu'une moindre activité hémodynamique au sein de ces aires cérébrales était associée à une moindre aversion pour l'atteinte à autrui (e.g., Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005).

PARTIE 1. CADRE THEORIQUE

CHAPITRE 1. SOUMISSION À L'AUTORITE

1.1.1. Paradigme d'obéissance de Milgram

Recrutement des participants, local et personnel

Des personnes volontaires ont été recrutées par voie de presse et voie postale pour participer à une étude sur la mémoire et l'apprentissage, moyennant une compensation financière (\$4.50 pour 1h de participation). Les études ont été réalisées dans un luxueux laboratoire de l'Université de Yale. Les participants étaient accueillis par l'expérimentateur, un homme d'une trentaine d'années portant une blouse grise, au « maintien impassible » et à l'expression « plutôt sévère » (Milgram, 1974). À leur arrivée, les participants se trouvaient en présence d'un autre volontaire, en réalité un compère de l'expérimentateur.

Méthode expérimentale

Dans un premier temps, l'expérimentateur expliquait au participant l'objectif de l'étude : tester par la méthode expérimentale les effets de la punition sur l'apprentissage. Un tirage au sort truqué permettait d'attribuer systématiquement le rôle du professeur au participant et le rôle de l'élève au compère. Professeur et élève étaient ensuite conduits dans une pièce où l'élève était alors installé sur une chaise munie de sangles. L'expérimentateur précisait que les sangles servaient à éviter tout mouvement excessif lors de la réception de chocs électriques. Il attachait ensuite une électrode sur le poignet de l'élève. Une crème était apposée pour éviter tout risque de lésion cutanée. Afin d'augmenter encore la crédibilité de la situation, l'expérimentateur précisait en réponse à une question de l'élève que « même si les chocs sont extrêmement douloureux, ils ne risquent pas de provoquer de lésion permanente » (Milgram, 1974). Le professeur était ensuite conduit dans une pièce contigüe par l'expérimentateur (*cf.* figure 1).

Tâche de mémorisation et test de rappel

L'élève devait dans un premier temps mémoriser une liste de paires de mots énoncées par le professeur à l'aide d'un micro (e.g., « ciel - bleu ») avant d'effectuer un test de rappel. Ce test de rappel constituait l'expérience proprement dite qui se déroulait comme suit : (i) pour chaque essai, le professeur devait énoncer l'un des termes de chaque paire suivi d'une liste de quatre mots (e.g., « bleu : compteur, ruban, ciel, yeux »), (ii) l'élève devait retrouver dans cette liste le terme initialement associé au premier mot et indiquer son choix à l'aide d'un boîtier réponse comprenant quatre boutons, (iii) le professeur était informé du choix à l'aide d'un panneau lumineux placé au-dessus du générateur de chocs et devait administrer un choc électrique en cas d'erreur.

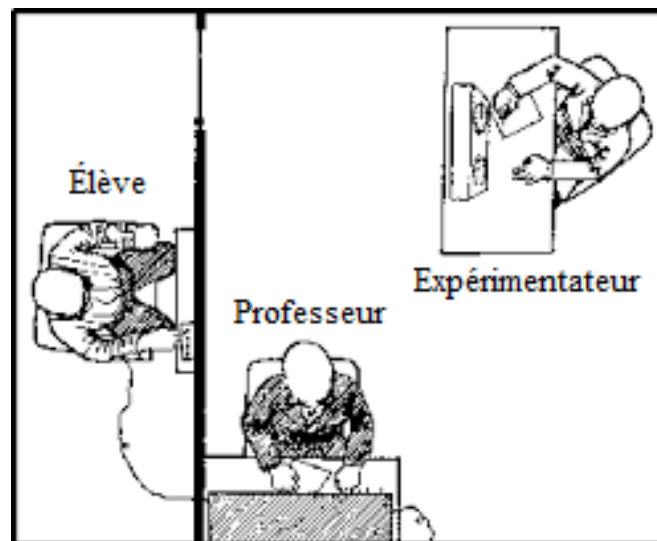


Figure 1. Disposition générale de l'espace expérimental. (Milgram, 1974)

Générateur de chocs électriques

Le professeur était installé face à un générateur de chocs électriques comportant trente manettes, chacune étant surmontée par l'indication d'un voltage. Ces indications allaient de 15 à 450 volts, avec une incrémentation de 15 volts (cf. figure 2). Chaque groupe de quatre manettes était associé à une mention : « Choc léger », « Choc modéré », « Choc fort », « Choc très fort », « Choc intense », « Choc extrêmement intense », « Attention choc dangereux ». Les deux dernières manettes portaient la mention « XXX ».

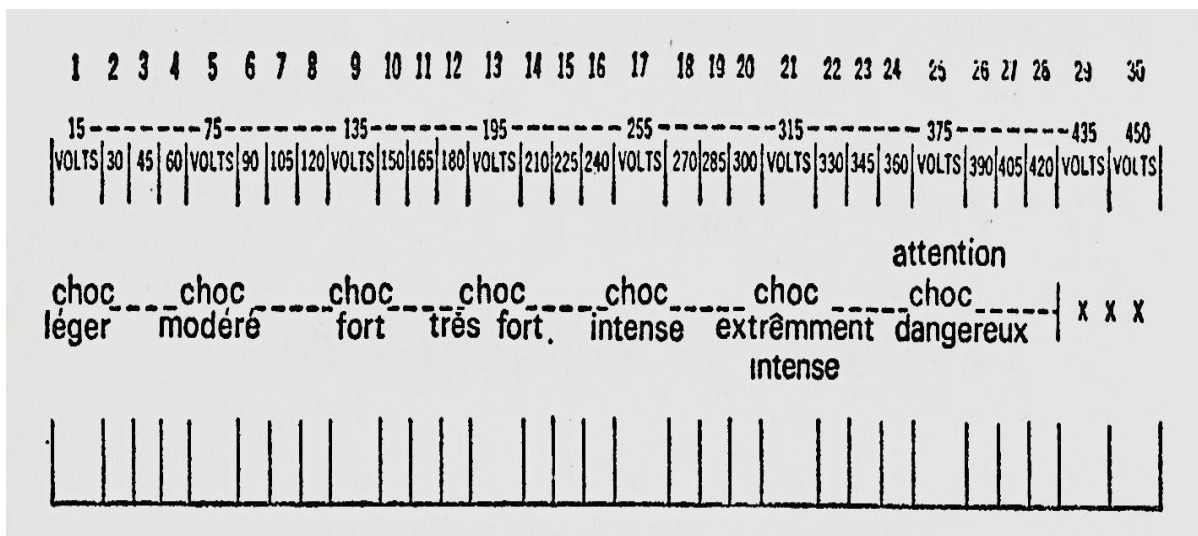


Figure 2. Représentation du générateur de chocs électriques. (Milgram, 1974)

Les instructions données aux participants précisait que non seulement un choc devait être administré à l'élève après chaque erreur, mais que l'intensité devait être augmentée de 15 volts à chaque nouvelle erreur. Afin de renforcer la crédibilité du dispositif, chaque participant se voyait administrer un choc témoin de 45 volts à l'aide de la troisième manette du générateur avant le début de l'expérience. L'expérience se terminait avec l'administration de trois chocs de 450 volts.

Injonctions de l'expérimentateur

En cas de refus de la part des participants de poursuivre l'expérience, l'expérimentateur formulait une série de quatre injonctions :

- (i) « *Veillez continuer, s'il vous plaît.* »
- (ii) « *L'expérience exige que vous continuiez.* »
- (iii) « *Il est absolument indispensable que vous continuiez.* »
- (iv) « *Vous n'avez pas le choix, vous devez continuer.* »

Ces injonctions étaient utilisées de manière séquentielle et l'expérience ne prenait fin qu'en cas de refus d'obéir à la quatrième injonction. La séquence complète était reprise à chaque nouvelle tentative du participant d'arrêter l'expérience. Le ton utilisé par l'expérimentateur était « ferme mais courtois » (Milgram, 1974).

Réactions de l'élève

Les réponses de l'élève étaient prédéterminées selon un *pattern* de trois réponses fausses pour une réponse juste. Dans l'expérience 1 (expérience princeps), le participant ne pouvait entendre les réactions de l'élève aux chocs électriques. Toutefois, à partir de 300 volts, le participant entendait l'élève bouger vivement sur sa chaise et marteler le sol. Suite à ces manifestations, l'élève ne fournissait plus aucune réponse. Dans l'expérience 2 (condition standard), le participant entendait les réactions de l'élève. Ces réactions étaient prédéterminées de la manière suivante : (i) aucun désagrément exprimé entre 15 et 75 volts, (ii) léger gémissement entre 90 et 105 volts, (iii) à 120 volts, l'élève crie à l'expérimentateur que les chocs deviennent douloureux, puis hurlement de douleur à 135 volts, (iv) à partir de 150 volts, l'élève retire son consentement et exige qu'on le laisse partir, (v) à 180 volts, l'élève crie qu'il ne supporte plus la douleur, et à 270 volts, l'élève pousse des cris d'agonie, (vi) à 300 volts, l'élève annonce d'une voix douloureuse et désespérée qu'il ne fournira plus

aucune réponse, (vii) après 330 volts, le participant n'entend plus l'élève. L'expérimentateur précisait que l'absence de réponse devait entraîner l'administration d'un choc.

Mesure de l'obéissance

La mesure retenue par Milgram est le choc maximum administré à la victime, de 0 (*i.e.*, refus d'administrer le moindre choc) à 30 (*i.e.*, 450 volts, pour les participants ayant administré la totalité des chocs).

Débriefing

Un débriefing était assuré à la fin de l'expérience, dont le but était à la fois de questionner le participant au sujet de sa perception de l'expérience mais également de fournir les informations et le soutien nécessaires à la prévention de toute atteinte sévère à l'estime de soi (*e.g.*, Elms, 1995). Dans un premier temps, les participants étaient informés du caractère fictif des chocs et de la bonne santé de l'élève, avant de rencontrer la personne ayant joué le rôle de l'élève. Un entretien était ensuite assuré afin d'évaluer l'état émotionnel des participants et leur rapport à l'étude. Enfin, les participants étaient invités à renseigner une série de questionnaires. Tous les participants étaient recontactés par voie postale afin d'être informés du détail de la méthode utilisée et des résultats obtenus. Un an après leur participation, un questionnaire complémentaire leur était adressé.

Prédictions comportementales

Milgram a demandé à un échantillon de psychiatres, d'étudiants et d'adultes issus de la population générale de prédire leur propre comportement dans la condition standard. Le détail du protocole leur a été présenté et un diagramme schématique du générateur de choc leur a été fourni. Chacune des personnes interrogées prévoit qu'elle désobéira au cours de l'expérience (*cf.* tableau 1).

Tableau 1
Prédictions du comportement adopté par soi dans la condition standard

Niveau de choc	Mention verbale et niveau de voltage	Psychiatres (n = 39)	Etudiants (n = 31)	Adultes (n = 40)
	Refus	2		3
	Choc léger			
1	15	1		
2	30			
3	45			1
4	60	1		1
	Choc modéré			
5	75	6	4	7
6	90	1	3	1
7	105	4		1
8	120	4	1	3
	Choc fort			
9	135	1	3	2
10	150	14	12	9
11	165		1	2
12	180	2	6	3
	Choc très fort			
13	195	2	1	1
14	210			
15	225			1
16	240			1
	Choc intense			
17	255			
18	270			
19	285			
20	300	1		3
	Choc extrêmement intense			
21	315			
22	330			
23	345			
24	360			
	Attention : choc dangereux			
25	375			
26	390			
27	405			
28	420			
	XXX			
29	435			
30	450			
Moyenne des niveaux de chocs maxima :		8.20	9.35	9.15
Pourcentage des prévisions de refus d'obéissance :		100%	100%	100%

Milgram a demandé à un deuxième échantillon (composition similaire) de prédire le comportement d'un groupe de personnes fictives. Les réponses convergent fortement : toutes les catégories de l'échantillon prévoient une désobéissance quasi unanime, à l'exception de quelques cas pathologiques (1% ou 2% de la population) qui seuls pourraient aller jusqu'à infliger un choc de 450 volts à une victime à l'agonie. Les réponses des psychiatres sont illustrées par la figure 3.

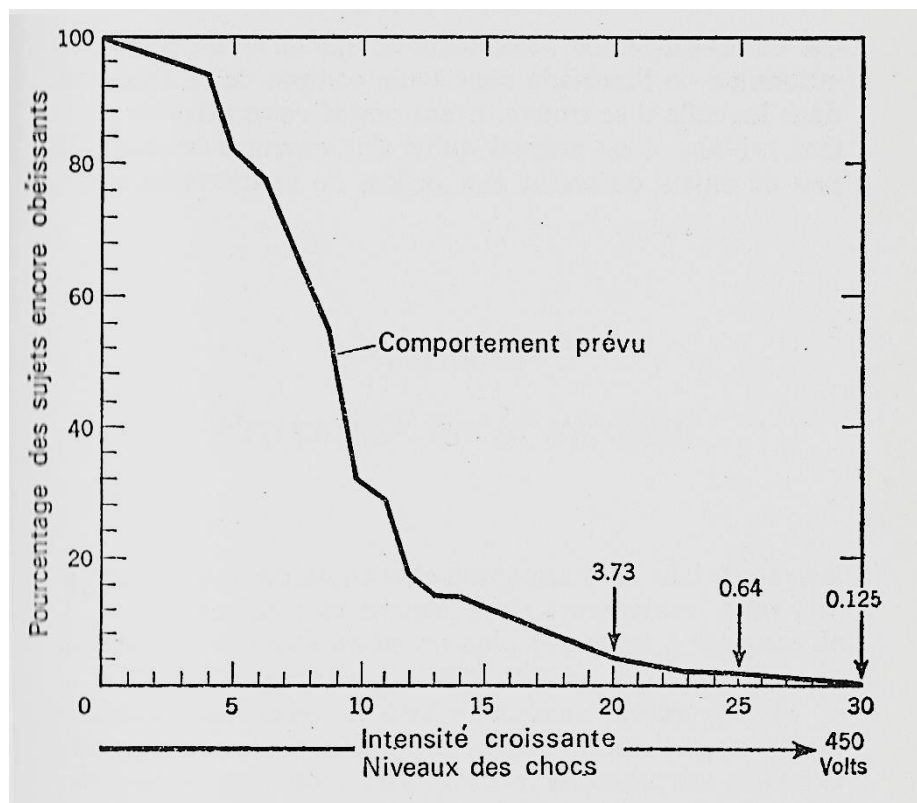


Figure 3. Prédications comportementales par un échantillon de psychiatres. (Milgram, 1974)

Selon les psychiatres, la majorité des participants devrait refuser d'aller au-delà du choc de 150 volts (premier retrait de consentement de la part de l'élève). Seulement 4% iraient jusqu'à administrer un choc de 300 volts, et 1 ou 2 participants sur 1000 pourraient infliger le choc le plus élevé prévu par le protocole (450 volts). Pour les personnes

interrogées, il semble que la coercition, la menace, ou un défaut d'empathie soient des conditions nécessaires pour augmenter l'OD (Milgram, 1974).

1.1.2. Résultats (expériences 1-4)

Les expériences 1-4 reposent sur la manipulation de la variable « proximité de la victime » : (i) dans l'expérience 1 (« Feedback à distance » – expérience princeps), professeur et élève ne se trouvent pas dans la même pièce et le professeur ne peut pas entendre l'élève, (ii) dans l'expérience 2 (« Feedback vocal » – condition standard), le professeur peut entendre l'élève, (iii) dans l'expérience 3 (« Proximité »), professeur et élève se trouvent dans la même pièce, (iv) dans l'expérience 4 (« Contact »), le professeur doit maintenir la main de l'élève sur une plaque délivrant les chocs électriques.

Les résultats sont illustrés par la figure 4, et sont présentés plus en détail dans le tableau 2. L'expérience 2 (« Feedback vocal ») demeure la plus connue : **62.5%** des participants ont obéi, en allant jusqu'à administrer trois chocs de 450 volts à une victime ayant sombré dans le silence après avoir poussé d'intenses cris de douleurs. Ce taux d'obéissance semble varier selon la proximité de la victime, celui-ci passant à 40% lorsque la victime se trouve dans la même pièce que le participant, et à 30% lorsque le participant doit maintenir la main de la victime. En revanche, lorsque les participants ne peuvent pas entendre la victime, le taux d'obéissance monte à 65%.

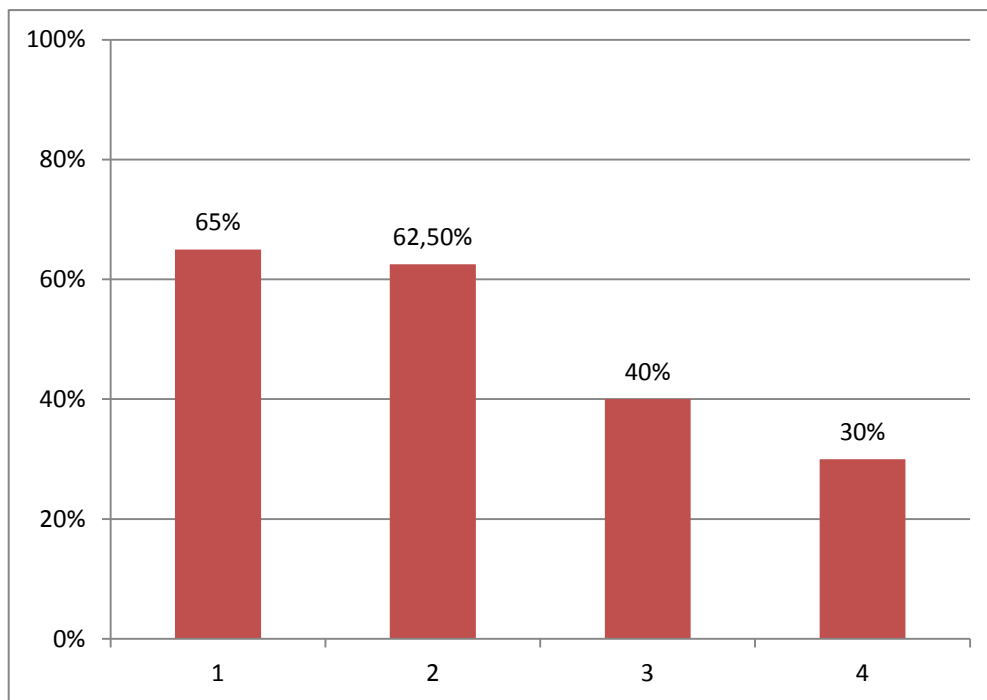


Figure 4. Pourcentages d'obéissance dans les expériences 1-4. (Milgram, 1974)

Tableau 2
Résultats des expériences 1-4 reposant sur la manipulation de la variable « proximité de la victime » (Milgram, 1974)

Niveau de choc	Mention verbale et niveau de voltage	Expérience 1 Sans feedback (n = 40)	Expérience 2 Feedback vocal (n = 40)	Expérience 3 Proximité (n = 40)	Expérience 4 Contact (n = 40)
	Choc léger				
1	15				
2	30				
3	45				
4	60				
	Choc modéré				
5	75				
6	90				
7	105			1	
8	120				
	Choc fort				
9	135		1		1
10	150		5	10	16
11	165		1		
12	180		1	2	3
	Choc très fort				
13	195				
14	210				1
15	225			1	1
16	240				
	Choc intense				
17	255			1	1
18	270				
19	285		1		1
20	300	5	1	5	1
	Choc extrêmement intense				
21	315	4	3	3	2
22	330	2			
23	345	1	1		1
24	360	1	1		
	Attention : choc dangereux				
25	375	1		1	
26	390				
27	405				
28	420				
	XXX				
29	435				
30	450	26	25	16	12
<hr/>					
Moyenne des niveaux de chocs maxima :		27	24.53	20.80	17.88
Pourcentage d'obéissance :		65%	62.5%	40%	30%

Réactions des participants

Plusieurs éléments suggèrent que les participants ont éprouvé un conflit psychologique souvent intense durant l'expérience. Tout d'abord, l'observation du comportement des participants durant l'expérience suggère l'expérience d'un stress modéré à intense (e.g., sudation, tremblements, tics nerveux, etc.). De plus, la formulation répétée d'injonctions à poursuivre s'est souvent avérée nécessaire pour obtenir un comportement d'obéissance (Milgram, 1974). Enfin, Milgram (1974) a réalisé une variante de la condition standard où le professeur devait décider du voltage à administrer à l'élève après chaque erreur, sans aucune autre directive de l'expérimentateur et sans formulation d'injonctions en cas de refus de poursuivre la procédure. Les résultats montrent que seulement 2.5% des participants ont administré la totalité des chocs.

Aspects sociodémographiques et interculturels

Les caractéristiques démographiques et/ou sociologiques des individus sont souvent considérées comme des déterminants importants du comportement en sciences humaines et sociales. Milgram s'est ainsi intéressé à l'influence de certains aspects comme le niveau d'éducation et la catégorie socioprofessionnelle. Alors que les modalités de recrutement ont permis à Milgram de constituer des échantillons hétérogènes du point de vue sociodémographique, les résultats ne montrent aucune influence du niveau d'éducation et de la catégorie socioprofessionnelle. Milgram (1974) n'ayant recruté que des hommes pour l'expérience 2, il a réalisé une variante de la condition standard où seules des femmes ont été recrutées. Le taux d'obéissance obtenu (65%) est similaire à celui observé dans l'expérience 2 (62.5%). Des hypothèses interculturelles ont été formulées (e.g., Mantell, 1971), mais les répliques exactes réalisées dans différents contextes culturels ne montrent aucune différence significative dans les taux d'obéissance (*cf.* tableau 3 ; e.g., Blass, 2012).

Tableau 3
Réplifications interculturelles des résultats de Milgram (condition standard)

Études	Pays	Taux d'obéissance (%)
Ancona & Pareyson (1968)	Italie	85%
Edwards, Franks, Friedgood, Lobbán, & Mackay (1969)	Afrique du Sud	87.5%
Mantell (1971)	Allemagne de l'Ouest	85%
Kilham & Mann (1974)	Australie	54%
Shanab & Yahya (1978)	Jordanie	62.5%
Miranda, Caballeor, Gomez, & Zamorano (1981)	Espagne	50%
Schurz (1985)	Autriche	80%

Les expériences 1-4 de Milgram s'inscrivent dans une série de 23 expériences portant essentiellement sur le rôle de facteurs situationnels. Il est nécessaire de souligner ici que le travail de recherche de Milgram ne s'est pas inscrit dans une démarche d'investigation systématique de facteurs articulés entre eux, mais plutôt dans une démarche d'exploration du rôle de facteurs souvent considérés isolément. Nous présenterons ci-dessous de manière synthétique l'ensemble des 23 expériences réalisées par Milgram (1963, 1965, 1974) ainsi que deux récentes méta-analyses des données pouvant contribuer à une meilleure compréhension des déterminants de l'OD.

1.1.3. Déterminants situationnels et dispositionnels de l'OD

1.1.3.1. Déterminants situationnels de l'OD

Le tableau 4 résume les 23 expériences réalisées par Milgram. Les résultats suggèrent que le taux d'obéissance serait bien affecté par certains facteurs situationnels, mais l'interprétation de ces résultats demeure peu aisée du fait (i) de l'hétérogénéité des facteurs manipulés, mais aussi (ii) de la taille des échantillons (entre 20 et 40 participants). La démarche la plus courante dans la littérature sur les travaux de Milgram consiste à effectuer des comparaisons deux à deux et à conclure sur cette seule base à l'influence de certains facteurs situationnels sur l'OD. Par exemple, il est courant de comparer la condition standard (62.5% d'obéissance) à la condition « Bridgeport » (47.5% d'obéissance) et de conclure que la légitimité influence l'OD. Cette démarche a motivé la formulation d'un certain nombre de propositions théoriques sur les déterminants de l'OD (e.g., Aronson, Wilson, & Akert, 2005 ; Blass, 1996 ; Gilovich, Keltner, & Nisbett, 2006 ; Kelman & Hamilton, 1989 ; Mixon, 1972, 1977 ; Myers, 2013 ; Nissani, 1990 ; Sutton & Douglas, 2013). Toutefois, certains auteurs ont eu recours aux méthodes statistiques afin de contribuer à une meilleure compréhension des résultats de Milgram autrement que par la seule interprétation spéculative.

Tableau 4
Description des 23 expériences réalisées par Milgram

N°	Condition	Brève description	Taux d'obéissance
1	<i>Sans feedback vocal</i> (n = 40)	Similaire à la condition standard, mais le professeur ne peut pas entendre l'élève.	65%
2	<i>Avec feedback vocal</i> (n = 40)	Condition standard , avec un professeur et un élève se trouvant dans deux pièces différentes, et un expérimentateur se trouvant dans la même pièce que le professeur.	62.5%
3	<i>Proximité avec l'élève</i> (n = 40)	Similaire à la condition standard, mais professeur et élève se trouvent dans la même pièce (le second est assis derrière le premier).	40%
4	<i>Contact avec l'élève</i> (n = 40)	Similaire à la condition standard, mais le professeur doit maintenir la main de l'élève sur une plaque délivrant les chocs électriques.	30%
5	<i>Troubles cardiaques</i> (n = 40)	Nouvelle condition standard . Similaire à la condition standard, mais l'élève signale une pathologie cardiaque au début de l'expérience, puis s'en plaint au cours de l'expérience.	65%
6	<i>Acteurs différents</i> (n = 40)	Similaire à la condition 5, mais des acteurs différents assurent les rôles de l'expérimentateur et de l'élève.	50%
7	<i>Pression groupale à la désobéissance</i> (n = 40)	Similaire à la condition standard, mais avec trois professeurs : deux compères et un participant. Les deux compères s'opposent à l'expérimentateur.	10%
8	<i>Conditions posées par l'élève</i> (n = 40)	Similaire à la condition standard, mais l'élève précise au début de l'étude qu'il ne donnera son consentement qu'à condition de pouvoir se retirer de l'expérience quand il le souhaite.	40%
9	<i>Pression groupale à l'obéissance</i> (n = 40)	Similaire à la condition 7, mais deux compères « professeurs » exercent une pression à l'obéissance sur le participant.	72.5%
10	<i>Instructions conflictuelles</i> (n = 20)	Similaire à la condition standard, mais l'expérimentateur exhorte le professeur à ne plus envoyer de chocs, alors que l'élève l'exhorte à poursuivre (obéir revient à ne pas aller jusqu'à 450 volts).	100%

Suite du Tableau 4

11	<i>Choix groupal</i> (n = 40)	Similaire à la condition 7, mais les professeurs peuvent déterminer le niveau du choc à administrer (le plus bas des trois choix): les compères professeurs se prononcent en premier et préconisent systématiquement une augmentation du voltage.	17.5%
12	<i>Permutation des rôles</i> (n = 20)	Similaire à la condition standard, mais l'expérimentateur et l'élève échangent leurs rôles (obéir revient à ne pas aller jusqu'à 450 volts).	100%
13	<i>Non administration des chocs</i> (n = 40)	Similaire à la condition 7, mais le professeur énonce les mots alors qu'un compère professeur administre les chocs.	92.0%
14	<i>Carte blanche</i> (n = 40)	Similaire à la condition standard, mais le professeur décide du voltage à administrer, sans aucune directive de l'expérimentateur.	2.5%
15	<i>Expérimentateur prosocial vs destructeur</i> (n = 20)	Similaire à la condition standard, mais deux expérimentateurs sont présents et formulent des demandes conflictuelles: l'un demande l'arrêt de l'envoi des chocs, l'autre la poursuite.	20%
16	<i>L'expérimentateur devient l'élève</i> (n = 20)	Similaire à la condition standard, mais deux expérimentateurs sont présents dont l'un prend la place de l'élève après que la personne initialement prévue se soit retirée de l'étude.	65%
17	<i>Compère professeur en charge</i> (n = 20)	Similaire à la condition standard, mais avec un compère professeur qui est chargé de décider du voltage à administrer alors que l'expérimentateur s'absente.	55%
18	<i>Absence de l'expérimentateur</i> (n = 40)	Similaire à la condition standard, mais l'expérimentateur s'absente et demande au professeur d'assurer le bon déroulement de l'expérience.	22.5%
19	<i>Autorité à distance</i> (n = 40)	Similaire à la condition 18, mais l'expérimentateur laisse des instructions préenregistrées que le professeur se doit de suivre.	37.5%

Suite du Tableau 4

20	<i>Femmes</i> (n = 40)	Similaire à la condition standard, mais les professeurs sont des femmes.	65%
21	<i>Autorité d'un pair</i> (n = 20)	Similaire à la condition 17, mais l'expérimentateur ne donne aucune consigne relative au choix du voltage.	20%
22	<i>Bridgeport</i> (n = 40)	Similaire à la condition standard, mais l'expérience se déroule dans un bureau modeste de Bridgeport et non à l'Université de Yale.	47.5%
23	<i>Proches dans le rôle de l'élève</i> (n = 20)	Similaire à la condition standard, mais l'élève est un proche du professeur.	15%

Méta-analyse de Haslam, Loughnan et Perry (2014)

Haslam, Loughnan et Perry (2014) ont réalisé une méta-analyse sur les données de 21 variantes, avec un échantillon total de 740 participants (les conditions 10 et 12 ont été exclues des analyses, la désobéissance dans ces variantes revenant à administrer la totalité des chocs). Préalablement à l'analyse, les auteurs ont proposé un regroupement des facteurs manipulés par Milgram en un nombre restreint de catégories. Une première approche « schématique » a conduit les auteurs à identifier *trois rôles* hiérarchiquement ordonnés (expérimentateur, professeur, élève) et *deux types de relations* (expérimentateur/professeur, professeur/élève). Suivant cette approche, les variantes introduites par Milgram concernent toutes l'un de ces cinq aspects, auxquels il faut ajouter le « *cadre* » (*i.e.*, lieu de l'expérimentation – Bridgeport vs Yale). Les auteurs ont ainsi défini *six classes* de facteurs manipulés par Milgram (*cf.* figure 5).

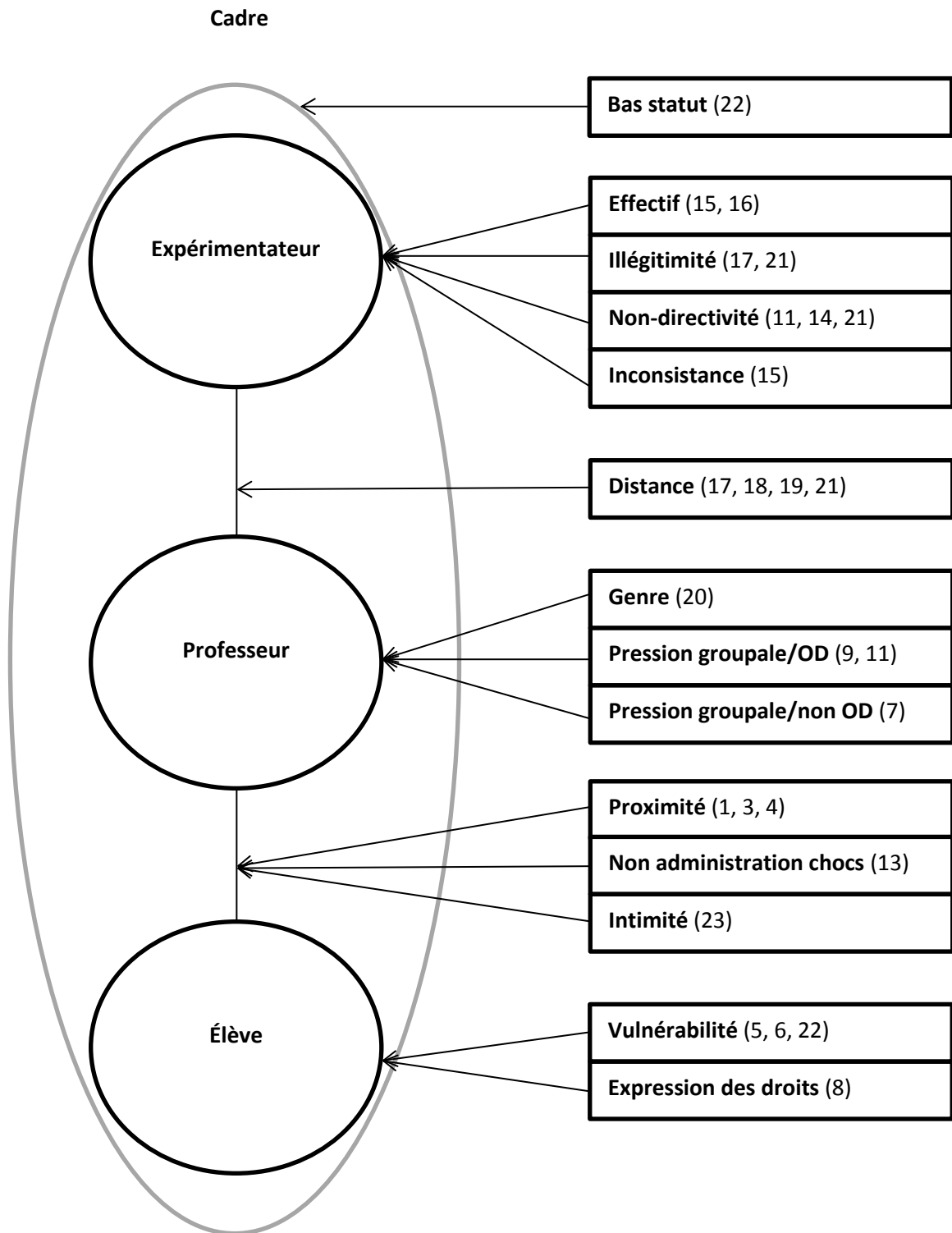


Figure 5. Représentation schématique des travaux de Milgram. (Haslam et al., 2014)

Les facteurs prédictifs de l'OD dans cette méta-analyse apparaissent relativement disparates (*cf.* tableau 5). Les variables associées aux effets statistiques les plus puissants, dans un ordre décroissant, sont (i) la directivité de l'expérimentateur, (ii) la pression groupale à la désobéissance, (iii) la proximité professeur/élève, et la non prise en charge de l'envoi des chocs, (iv) la distance professeur/expérimentateur, (v) la relation d'intimité professeur/élève, et (vi) l'inconsistance de l'expérimentateur. Certains de ces effets ont été repérés dans la littérature sur Milgram, comme la proximité, la pression groupale et la distance entre le professeur et l'expérimentateur. En revanche, certaines variables comme la relation d'intimité professeur/élève ne sont pas citées dans la littérature sur les travaux de Milgram (Haslam et al., 2014).

Tableau 5
Résultats de la méta-analyse de Haslam et al. (2014)

Classes de facteurs	Variables affectant le taux d'obéissance
Relation professeur/élève	Les trois variables de cette classe affectent le taux d'obéissance : (i) l'augmentation de la proximité avec l'élève diminue le taux d'obéissance, (ii) l'existence d'une relation d'intimité entre professeur et élève est associée à un moindre taux d'obéissance, (iii) la non prise en charge de l'envoi des chocs par les professeurs augmente le taux d'obéissance.
Relation professeur/expérimentateur	Plus la distance augmente entre professeur et expérimentateur, plus le taux d'obéissance diminue.
Caractéristiques de l'expérimentateur	La directivité augmente l'obéissance alors que l'inconsistance la diminue.
Caractéristiques du professeur	La pression groupale à la désobéissance diminue le taux d'obéissance.
Caractéristiques de l'élève	Les caractéristiques associées à l'élève n'influencent pas le taux d'obéissance.
Cadre	Le lieu de l'expérimentation n'affecte pas le taux d'obéissance.

Le fait que les participants ne se montrent pas moins obéissants quand l'autorité est exercée par un pair, de même que l'absence d'effet du lieu d'expérimentation, ont conduit Haslam et al. (2014) à proposer que les effets de la légitimité aient été surévalués dans la littérature sur les travaux de Milgram (e.g., Kelman & Hamilton, 1989 ; Passini & Morselli, 2009). En revanche, il ressort de cette méta-analyse que toutes les variables caractérisant la relation professeur/élève affectent significativement le taux d'obéissance.

Méta-analyse de Packer (2008)

La méta-analyse de Packer (2008) semble également soutenir l'idée d'un rôle important de la relation professeur/élève dans le paradigme de Milgram. L'auteur a réalisé une méta-analyse de 8 expériences de Milgram (variantes 2-6, 8-10) dans le but de dégager certains « invariants » dans le comportement de désobéissance. Les résultats suggèrent que la désobéissance se produit essentiellement à 150 volts, moment où la victime demande pour la première fois à être libérée. Il ressort également que le taux d'obéissance dans les différentes études covarie avec le taux de désobéissance à 150 volts : plus l'obéissance diminue, plus la désobéissance à 150 volts augmente. Selon Packer (2008), ces résultats mettent en évidence un point critique de décision dans le paradigme de Milgram et suggèrent que les participants désobéissants perçoivent le droit de la victime à arrêter l'expérience comme supplantant les ordres de l'expérimentateur.

Si ces résultats suggèrent l'influence de certaines variables situationnelles, ils montrent également l'existence de fortes différences interindividuelles dans l'OD : (i) un peu plus d'un tiers des participants a désobéi dans la condition standard (Milgram, 1963), (ii) dans toutes les expériences réalisées, on observe la présence de participants déviants relativement au *pattern* comportemental attendu (e.g., Blass, 1991). De plus, la taille d'effet des variables comme la proximité de la victime et de l'expérimentateur est statistiquement faible (convertie

en coefficient de corrélation, elle se situe entre .30 et .40 ; voir Funder, Guillaume, Kumagai, Kawamoto, & Sato, 2012). Bien que l'importance des différences interindividuelles dans les résultats de Milgram ait souvent été sous-estimée (e.g., Bègue et al., 2015 ; Berkowitz, 1999 ; Blass, 1991), des auteurs ont exploré l'influence de variables dispositionnelles théoriquement liées à l'OD. Nous présenterons dans un premier temps les travaux relatifs à l'influence des attitudes et de la personnalité sur l'OD, avant de présenter les travaux portant sur le rôle des compétences socio-émotionnelles.

1.1.3.2. Déterminants dispositionnels de l'OD

Il nous faut signaler ici que la recherche relative à l'influence de variables dispositionnelles sur l'OD présente un certain nombre de problèmes :

Problèmes méthodologiques

Il existe d'importantes différences inter-études dans les procédures expérimentales utilisées. Par exemple, dans l'étude de Haas (1966), les participants étaient des membres du personnel de la direction d'une entreprise qui se voyaient ordonner par la direction de formuler des évaluations critiques de leurs supérieurs et de préciser lesquels devaient être licenciés. Ces recommandations étaient réparties en six classes, allant du refus de participer jusqu'à la recommandation de licenciement. Dans l'étude de Miller (1975), les participants devaient s'auto-infliger des chocs électriques (en se saisissant de fils électriques sous tension) pendant qu'ils réalisaient une tâche de résolution de problèmes arithmétiques. Ces différences inter-études empêchent bien souvent la mise en relation des résultats obtenus.

Problèmes psychométriques

Bon nombre d'études réalisées dans les années 1960 et 1970 reposent sur l'usage de questionnaires aux qualités psychométriques insuffisantes, comme le « Social Intelligence

Test » (Moss & Hunt, 1927) utilisé dans l'étude de Burley et McGuiness (1977). L'exploitation de leurs résultats est de ce fait impossible.

Ainsi, nous avons systématiquement exclu de la présente revue de littérature les études cumulant (i) l'usage de questionnaires aux qualités psychométriques insuffisantes et (ii) le recours à un protocole expérimental n'impliquant pas une atteinte incrémentale à l'intégrité physique d'autrui (ou de soi). Les études retenues sont présentées dans le tableau 6.

Tableau 6

Protocole expérimental et principales mesures des études relatives aux déterminants dispositionnels de l'OD

Études	Variabes	Protocoles expérimentaux	Échelles
Elms & Milgram (1966)	Autoritarisme	Les auteurs ont recontacté une partie des participants aux expériences 1-4 pour leur faire remplir une batterie de questionnaires.	Échelle F (Adorno et al., 1950)
Bock & Warren (1972)	Religiosité	Variante 5 du paradigme de Milgram.	Échelle de mesure de l'orientation religieuse (King & Hunt, 1972) - Echelle d'orientation religieuse (Allport & Ross, 1967)
Miller (1975)	Autoritarisme	Les sujets devaient s'auto-infliger des chocs électriques (en se saisissant de fils électriques sous tension) pendant une tâche de résolution de problèmes arithmétiques.	Échelle F (Adorno et al., 1950)
Altemeyer (1981)	Autoritarisme de droite	Protocole similaire à celui de Miller (1975).	Échelle d'autoritarisme de droite (Altemeyer, 1981)
Burger (2009)	Souci empathique	Réplication partielle de la variante 5 (arrêt de l'expérience à 150 volts).	Indice de réactivité interpersonnelle (Davis, 1983)
Dambrun & Vatiné (2010)	Autoritarisme de droite	« Immersive Video Milgram Obedience Experiment » : reproduit la variante 5, avec feedback audiovisuel.	Échelle d'autoritarisme de droite (Altemeyer, 1988)
Zeigler-Hill, Southard, Archer, & Donohoe (2013)	Neuroticisme - Émotions négatives	Administration de chocs sonores à un élève souffrant de migraines.	« Big-Five Inventory » (John & Srivastava, 1999) - PANAS (« Positive and Negative Affect Schedule » ; Watson, Clark, & Tellegen, 1988)
Bègue et al. (2015)	Caractère consciencieux, agréabilité - Orientation politique	« Jeu de la mort » : transposition du paradigme de Milgram dans le contexte d'un faux jeu télévisé.	« Big-Five » Mini-Marqueurs questionnaire (Saucier, 1994) - Item de placement politique (1 = « Extrême gauche » ; 10 = « Extrême droite »)

1.1.3.2.1. Influence de la personnalité et des attitudes sur l'OD

La recherche sur l'influence de variables dispositionnelles théoriquement reliées à l'OD a été initiée par Elms et Milgram (1966), qui se sont intéressés au rôle de l'*autoritarisme* (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). Elms et Milgram (1966) ont recontacté une partie des participants aux expériences 1-4 afin de leur faire remplir une batterie de questionnaires incluant l'échelle F (Adorno et al., 1950). Les *items* de cette échelle sont organisés en neuf catégories : (i) conventionnalisme (adhésion rigide aux conventions sociales), (ii) soumission autoritaire (attitude de soumission non critique à l'égard des autorités traditionnelles), (iii) agression autoritaire (tendance punitive à l'encontre des personnes transgressant des normes conventionnelles), (iv) anti-intracception (opposition aux personnes imaginatives et compatissantes), (v) superstition et stéréotypie (mysticisme et rigidité mentale), (vi) puissance et rudesse (attrait pour l'exercice du pouvoir et mentalité dure), (vii) caractère destructif et cynisme (hostilité généralisée, dénigrement de l'humain), (viii) projectivité (disposition à croire que le monde est dangereux, projection d'impulsions émotionnelles inconscientes), (ix) sexe (souci exagéré pour les conduites sexuelles). Les résultats montrent que les participants obéissants ont des scores supérieurs à l'échelle F, suggérant ainsi que l'autoritarisme serait un prédicteur de l'OD (voir aussi Elms, 1972).

Bock (1972) a examiné l'influence de la *religiosité* sur l'OD en utilisant la variante 5 du paradigme de Milgram (condition standard avec un élève souffrant de problèmes cardiaques). L'identité de l'expérimentateur a également été manipulée (autorité scientifique vs religieuse vs neutre). Les résultats montrent que plus les participants ont des scores élevés sur divers questionnaires d'orientation religieuse (religion chrétienne) plus ils tendent à obéir face à une autorité traditionnelle (*i.e.*, scientifique ou religieuse). À l'inverse, les participants

ayant les scores de religiosité les plus faibles tendent à désobéir à toutes les formes d'autorité (*i.e.*, scientifique, religieuse ou neutre ; voir Blass, 1991).

Miller (1975) et Altemeyer (1981) ont utilisé un paradigme d'obéissance impliquant non plus une atteinte à autrui mais à soi. Ainsi, dans l'étude de Miller (1975), les participants devaient s'auto-infliger des chocs électriques (en se saisissant de fils électriques sous tension) pendant une tâche de résolution de problèmes arithmétiques. L'auteur a identifié une relation significative et positive entre les scores des participants à l'échelle F et le niveau d'obéissance. Les faiblesses psychométriques de l'échelle F ont toutefois conduit à son abandon, au profit de l'échelle d'*autoritarisme de droite* développée par Altemeyer (1981). En utilisant un protocole expérimental similaire à celui de Miller (1975), Altemeyer (1981) a examiné la relation entre les scores des participants à son échelle d'autoritarisme et leur niveau d'obéissance. Les résultats montrent une relation significative et positive : plus les participants ont un score élevé d'autoritarisme de droite, plus ils tendent à obéir. Beaucoup plus récemment, Dambrun et Vatiné (2010) ont trouvé que les scores à l'échelle d'autoritarisme de droite des participants prédisaient leur niveau d'obéissance dans un protocole vidéo immersif reproduisant la variante 5 (« Immersive Video Milgram Obedience Experiment »).

Suivant une procédure similaire à celle d'Elms et Milgram (1966), Bègue et al. (2015) ont recontacté les participants au « Jeu de la mort », une transposition du paradigme de Milgram dans le contexte d'un faux jeu télévisé (voir Beauvois et al., 2012, pour une présentation détaillée du protocole). Bègue et al. (2015) se sont intéressés à l'influence de deux traits de personnalité particulièrement susceptibles selon eux de prédire l'OD : le *caractère consciencieux* et l'*agréabilité* (e.g., McCrae & Costa, 1987). Ces deux traits sont souvent considérés comme des composantes centrales de la conformité sociale (e.g.,

DeYoung, Peterson, & Higgins, 2002 ; Duckitt & Sibley, 2010). Les auteurs se sont également intéressés à l'influence de l'*orientation politique* des participants, la littérature montrant que les valeurs associées à la gauche politique sont négativement reliées à la conformité sociale (e.g., Schwartz, Caprara, & Vecchione, 2010) et à l'autoritarisme de droite (e.g., Duckitt & Sibley, 2010). Les résultats montrent que l'agréabilité et le caractère consciencieux sont positivement reliés à l'OD, de même que le placement sur la droite du spectre politique (*i.e.*, conservatisme politique).

Ces études suggèrent un rôle central de l'autoritarisme dans l'OD : (i) trois études montrent de manière consistante une influence de l'autoritarisme sur l'OD, (ii) les travaux sur l'autoritarisme suggèrent que les autres prédicteurs de l'OD (religiosité, conformité sociale et conservatisme politique) sont fortement reliés à l'autoritarisme (e.g., Altemeyer, 1996 ; Sibley & Duckitt, 2008). Si l'état actuel des connaissances sur la soumission à l'autorité ne permet pas d'expliquer par quels mécanismes l'autoritarisme influence l'OD, les travaux réalisés en psychologie politique sur les antécédents de l'autoritarisme peuvent aider à comprendre cette influence.

1.1.3.2.2. Autoritarisme

Premières propositions théoriques

Adorno et al. (1950) ont initié la recherche empirique sur l'autoritarisme, et leurs résultats montrent que l'hostilité intergroupe, la valorisation ethnocentrique du groupe d'appartenance, le conservatisme politico-économique et les attitudes profascistes covarient pour former un « syndrome attitudinal général ». Selon les auteurs, ce syndrome serait l'expression d'une dimension de la personnalité regroupant neuf traits covariants (la « personnalité autoritaire ») que l'on pourrait mesurer avec leur « échelle F ». Adoptant une perspective freudo-marxiste, les auteurs ont proposé que la personnalité autoritaire soit

associée à une « vulnérabilité émotionnelle » héritée d'une éducation parentale « punitive ». Ces premiers travaux sur l'autoritarisme ont suscité un grand intérêt mais également un certain nombre de critiques : (i) la perspective psychanalytique adoptée ne bénéficie d'aucune validation empirique, (ii) l'échelle F présente d'importantes faiblesses psychométriques (e.g., manque de fiabilité et d'unidimensionnalité, l'échelle F mesurant en fait plusieurs traits faiblement reliés ; Altemeyer, 1981).

Des auteurs comme Allport (1954), Rokeach (1954) et Wilson (1973) ont proposé une conceptualisation plus soignée de la personnalité autoritaire. Ainsi, Allport (1954) a suggéré que l'autoritarisme se caractérise par une vulnérabilité aux *sentiments d'insécurité et de peur*. Ces sentiments favoriseraient chez les personnes autoritaires un besoin de structure, d'ordre et de contrôle, ainsi que des réactions punitives hostiles en réponse au contre-normatif, au changement et à la nouveauté. Rokeach (1954, 1960) a également proposé une conceptualisation simplifiée de la personnalité autoritaire avec le concept de « dogmatisme » défini comme « une *organisation cognitive relativement close* de croyances et non-croyances relative à la réalité ». Cette configuration cognitive serait à l'origine d'une difficulté à traiter les informations non congruentes avec les croyances acquises, et prédisposerait à l'autoritarisme en général (et non seulement de droite). Elle motiverait également l'hostilité et le rejet envers les personnes ou groupes porteurs de croyances et valeurs dissimilaires. Wilson (1973) a quant à lui proposé comme alternative au concept d'autoritarisme celui de conservatisme, défini comme « une susceptibilité généralisée à faire *l'expérience de la peur ou de l'anxiété* en réponse à l'incertitude ». Rokeach (1954) et Wilson (1973) ont chacun élaboré une échelle de mesure présentant d'importantes faiblesses psychométriques (e.g., faible consistance interne ; Altemeyer, 1981, 1996).

Altemeyer et l'échelle d'autoritarisme de droite

L'échelle d'autoritarisme de droite (« right-wing authoritarianism » : RWA) développée par Altemeyer (1981, 1988, 1996) constitue la première mesure fiable et unidimensionnelle de l'autoritarisme. Comme ses prédécesseurs, Altemeyer conceptualise le RWA comme l'expression d'une dimension de la personnalité. Les *items* de l'échelle RWA sont réunis en trois catégories : (i) soumission autoritaire, (ii) agression autoritaire, (iii) conventionnalisme. L'auteur a réalisé un important travail de validation de son échelle, que nous présentons brièvement ci-dessous.

Soumission autoritaire. Un grand nombre de recherches supporte l'idée voulant que l'échelle RWA mesure en partie un trait exprimant la soumission autoritaire. Par exemple, les scores à l'échelle RWA corrént positivement avec la légitimation des actions gouvernementales illégales dirigées contre des groupes « non conventionnels », incluant différentes formes d'atteintes aux droits et libertés de cibles comme les dissidents politiques (Altemeyer, 1981).

Agression autoritaire. Les scores à l'échelle RWA corrént également positivement avec la sévérité du traitement pénal préconisé à l'encontre des personnes responsables d'une violation de la loi (à l'exception des figures d'autorité ; Altemeyer, 1981, 1988). Il est à noter que le RWA n'est associé qu'aux formes d'agression servant le renforcement des lois et valeurs.

Conventionnalisme. Enfin, les scores à l'échelle RWA corrént également positivement avec la religiosité, notamment le fondamentalisme religieux (défini comme la manifestation religieuse de l'autoritarisme ; Altemeyer, 1996), ainsi qu'avec les rôles traditionnels de genre, le conservatisme politique et le traditionalisme (Altemeyer, 1996).

Recherche récente sur les origines du RWA

La recherche sur les origines du RWA a majoritairement porté sur le rôle de facteurs environnementaux comme l'apprentissage social et l'environnement familial (e.g., Duckitt, 2001 ; Hatemi & McDermott, 2012). Ce choix s'inscrit dans un débat historique au sein de la culture occidentale portant sur l'importance relative de la génétique et de l'environnement dans la détermination des traits humains (nature vs culture). L'étude des traits sociaux s'est longtemps inscrite dans un paradigme voulant que les différences interindividuelles soient socialement déterminées, et que l'humain se distinguerait des autres espèces animales pour avoir dépassé son héritage évolutif (e.g., Lumsden & Wilson, 1981). Partant de cette transcendance, les approches courantes en sciences sociales stipulent que la transmission intergénérationnelle des traits politiques serait la conséquence unique de mécanismes sociaux. Si nature et culture sont encore souvent considérées comme des forces distinctes et antagonistes (e.g., Charney, 2008), de nombreuses études tendent à démontrer le contraire (*cf.* théorie de la coévolution gène-culture ; Boyd & Richerson, 1985 ; Cavalli-Sforza & Feldman, 1981 ; Chiao & Blizinsky, 2009 ; Rushton, Littlefield, & Lumsden, 1986).

Plusieurs études de jumeaux ont rapporté de fortes corrélations entre les scores de RWA de jumeaux monozygotes et dizygotes élevés séparément (des résultats similaires ont été obtenus avec des mesures étroitement associées au RWA comme le traditionalisme, le conservatisme et la religiosité ; McCourt, Bouchard, Lykken, Tellegeen, & Keyes, 1999 ; Stöbell, Kämpfe, & Riemann, 2006). De plus, les résultats de ces études ont montré des corrélations supérieures chez les jumeaux monozygotes (vs dizygotes) élevés séparément. À ce jour, les travaux réalisés suggèrent une héritabilité de l'ordre de 60%, magnitude similaire à celle trouvée pour les mesures de personnalité comme le « Big-Five » (e.g., Bouchard, 2009 ; Koenig & Bouchard, 2006).

Des études ont également montré un effet de la menace sociale (e.g., crises économiques et politiques) : (i) la perception d'une menace augmente le RWA tant au niveau sociétal (Doty, Winter, Peterson, & Kimmelmeier, 1997 ; McCann, 1997 ; Sales, 1972) qu'individuel (Rickert, 1998 ; Rohan & Zanna, 1996 ; Sales & Friend, 1973), (ii) les personnes présentant un fort score de RWA sont plus réactives à la menace (Feldman & Stenner, 1997 ; Greenberg et al., 1990 ; Lavine et al., 1999 ; Lavine, Lodge, & Freitas, 2005), (iii) la manipulation expérimentale de la saillance de la menace augmente mécaniquement le niveau de RWA (Altemeyer, 1996 ; Duckitt, 2013 ; Duckitt & Fisher, 2003 ; Jugert & Hiemisch, 2005).

D'autres travaux ont mis en avant une relation positive entre certains traits de personnalité et le RWA. Des études ont montré de manière consistante une relation positive entre RWA et caractère consciencieux ainsi qu'une relation négative entre ouverture à l'expérience et RWA (voir la revue de Sibley & Duckitt, 2008). D'autres auteurs ont trouvé une relation positive entre RWA et divers traits exprimant des besoins d'ordre, de clôture et de structure (Cornelis & Van Hiel, 2006 ; Kossowska & Van Hiel, 2003 ; Van Hiel et al., 2004).

Orientation à la dominance sociale

Le concept d'orientation à la dominance sociale (« social dominance orientation » : SDO) a été proposé dans les années 1990 et est défini comme une orientation attitudinale exprimant une préférence générale pour des relations intergroupes égalitaires ou hiérarchiques (Pratto, Sidanius, Stallworth, & Malle, 1994 ; Sidanius & Pratto, 1999). Des études ont établi que l'échelle SDO prédit un ensemble de phénomènes autoritaires similaires à ceux prédis par l'échelle RWA (e.g., intolérance généralisée, préférence pour les partis politiques de droite, nationalisme, patriotisme, militarisme, support pour la peine de mort ; Pratto, Sidanius, & Levin,

2006 ; Sidanius & Pratto, 1999). Cependant, la recherche suggère que ces échelles mesurent des dimensions différentes et relativement indépendantes de l'autoritarisme (Altemeyer, 1998 ; Duckitt, 2001). Tout d'abord, le contenu des *items* des deux échelles diffère : (i) les *items* de l'échelle RWA expriment des croyances relatives au contrôle social, au respect et à l'obéissance en direction des autorités traditionnelles, à la conformation aux normes et valeurs religieuses, (ii) les *items* de l'échelle SDO expriment le soutien aux inégalités socioéconomiques et au droit des groupes dominants à exercer leur domination sur les groupes dominés. Ensuite, les scores de RWA et de SDO ne corréleront pas toujours avec les mêmes variables : seul le RWA est corrélé à la religiosité et à la valorisation de l'ordre, la structure, la conformité et la tradition, alors que seule la SDO est corrélée à la valorisation du pouvoir, de la réalisation de soi, de l'hédonisme et à la dominance (Altemeyer, 1998 ; Duckitt, 2001 ; Ekehammar & Akrami, 2003 ; McFarland, 2005 ; Van Hiel & Mervielde, 2002). De plus, seul le RWA est influencé par la menace sociale et est corrélé à la perception d'un environnement social dangereux, alors que seule la SDO est associée à un certain « darwinisme social » consistant à percevoir le monde comme un univers hautement compétitif où « les forts gagnent et les faibles perdent » (e.g., Duckitt & Sibley, 2010). Partant de l'ensemble de ces données, Altemeyer (1998) a proposé que SDO et RWA soient des dimensions indépendantes mais complémentaires de la personnalité autoritaire, l'une orientée vers l'exercice du pouvoir et l'autre orientée vers l'obéissance. La confusion entre ces deux dimensions pourrait en partie expliquer les difficultés tant théoriques que méthodologiques rencontrées avant les travaux d'Altemeyer (1981). Par exemple, l'échelle F comporte des *items* couvrant des aspects liés à la dominance et d'autres à l'obéissance.

Nouvelles perspectives sur l'autoritarisme

Des auteurs ont proposé dès la fin des années 1980 que les *items* des échelles RWA et SDO expriment non pas des dispositions comportementales ou des traits (comme cela est le

cas pour les échelles de personnalité) mais plutôt des croyances et attitudes de nature idéologique (Duckitt, 1989, 2001 ; Feldman & Stenner, 1997 ; Rosier & Willig, 2002 ; Saucier, 2000 ; Stone, Lederer, & Christie, 1993). Dans cette perspective, RWA et SDO seraient non pas l'expression d'une dimension de la personnalité mais plutôt d'*attitudes idéologiques* endossées pour satisfaire certaines *motivations ou besoins* (e.g., Duckitt, 1989, 2001)¹. Cette proposition est cohérente avec les travaux montrant que les attitudes idéologiques sont sensibles aux manipulations expérimentales et aux changements sociopolitiques (Duckitt & Fisher, 2003 ; Guimond, Dambrun, Michinov, & Duarte, 2003 ; Huang & Liu, 2005 ; Schmitt, Branscombe, & Kappen, 2003 ; Sales, 1973 ; Sales & Friend, 1973).

Modèle duel motivationnel

Le modèle duel motivationnel (e.g., Duckitt & Sibley, 2010) intègre plusieurs facteurs situationnels et dispositionnels dans le but d'expliquer à la fois les antécédents et les conséquences des attitudes idéologiques. Si RWA et SDO prédisent ensemble un certain nombre de phénomènes autoritaires, le modèle duel motivationnel propose que RWA et SDO soient des dimensions distinctes de l'autoritarisme exprimant chacune des buts et valeurs motivés rendus chroniquement saillants chez les individus par leur personnalité et leurs visions du monde (e.g., Duckitt, 2001). Ce modèle prévoit trois classes d'antécédents aux attitudes idéologiques : la personnalité, l'environnement social et les motivations.

Motivation. Un fort RWA exprimerait le but motivé à l'établissement et au maintien de la sécurité collective (*i.e.*, ordre social, cohésion, stabilité), but rendu chroniquement saillant par

¹ L'approche de la « cognition sociale motivée » propose qu'un individu puisse endosser des valeurs et croyances pour satisfaire certains besoins. En plus des motivations basées sur le métabolisme vital et la reproduction, la psychologie sociale a permis la définition d'un certain nombre de motivations : motivation au contrôle, motivations épistémiques, motivation à la protection du Soi, motivation d'affiliation (e.g., Jost, Glaser, Kruglanski, & Sulloway, 2003).

une vision du monde comme intrinsèquement dangereux et menaçant. Une forte SDO exprime le but motivé à l'inégalité et à la dominance, but rendu chroniquement saillant par une vision du monde comme « une jungle impitoyable et compétitive » où les forts gagnent et les faibles perdent.

Personnalité. La conformité sociale (faible ouverture d'esprit et fort caractère consciencieux dans le cadre du modèle du « Big Five » ; e.g., McCrae & John, 1992) prédisposerait au RWA, ce facteur conduisant les personnes à valoriser l'ordre, la stabilité, la sécurité et influençant les croyances relatives à la dangerosité du monde. Par contraste, une « mentalité rude » (faible agréabilité dans le cadre du « Big Five ») prédisposerait à la SDO, type de personnalité associé à une perception du monde comme un environnement « impitoyable et compétitif ».

Environnement social. Un environnement menaçant favoriserait le RWA *via* le développement d'une perception chronique et généralisée de la menace, alors qu'un environnement compétitif (marqué par les inégalités) favoriserait la SDO *via* une vision du monde « darwiniste sociale ».

Ces hypothèses sur les origines des attitudes idéologiques sont confortées par des études corrélationnelles utilisant la modélisation en équation structurelle et par des études longitudinales montrant l'effet causal de ces profils et visions du monde sur les scores aux échelles RWA et SDO. Elles sont également soutenues par des études expérimentales reposant sur la manipulation de la saillance de certains groupes ou environnements et montrant un effet causal des environnements décrits dans le modèle duel motivationnel tant sur les visions du monde que sur les scores aux échelles RWA et SDO (*cf.* figure 6 ; pour une revue, voir Duckitt & Sibley, 2010 ; voir aussi Sibley & Duckitt, 2008).

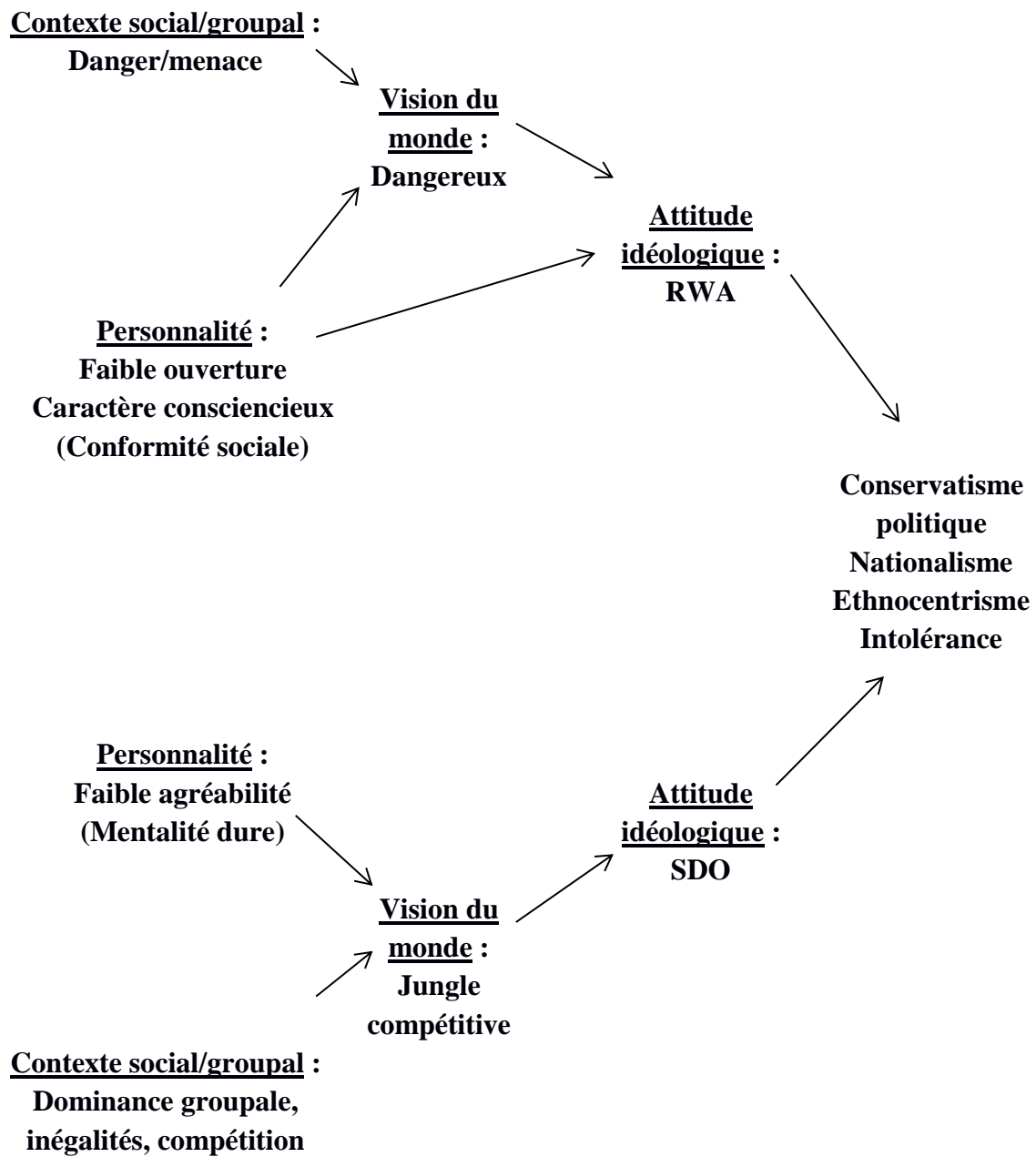


Figure 6. Modèle duel motivationnel (Duckitt, 2001 ; Duckitt & Sibley, 2010) proposant que la personnalité et l'environnement social influencent les croyances relatives aux caractéristiques du monde social, croyances venant à leur tour influencer les attitudes idéologiques. Ces attitudes détermineront les comportements sociopolitiques. (Adapté de Duckitt & Sibley, 2010)

Modèle de la sécurité collective

Jugert et Duckitt (2009) ont présenté le concept de « motivation à la sécurité collective » (MSC) comme facteur central dans la genèse du RWA. La MSC est définie comme la motivation pour l'individu de voir son collectif d'appartenance (groupe, société) bénéficier d'un environnement sûr, prédictible, harmonieux, stable, cohésif et ordonné. Dans le modèle proposé par les auteurs, la MSC médiatise les effets de la personnalité et des croyances en un monde dangereux sur le RWA (*cf.* figure 7).

Les personnes possédant une forte MSC seront particulièrement motivées à l'établissement, au maintien et à la défense de la sécurité collective au sein de leur groupe. Les traits de personnalité prédisant le RWA exprimant des besoins de structure, d'ordre et de clôture, la MSC peut être appréhendée comme l'expression de ces besoins dans la sphère sociale au travers de l'endossement de buts et valeurs. Ces besoins et préférences pour l'ordre et la structure rendront chroniquement saillante la motivation à l'établissement d'un ordre social stable, structuré et cohésif.

Jugert et Duckitt (2009) ont proposé que les trois dimensions du RWA défini par Altemeyer (1981) puissent être l'expression de différentes facettes de la MSC : (i) besoin de cohésion, consensus, conformité et harmonie (« *soumission autoritaire* »), (ii) contrôle social et ordre (« *agression autoritaire* »), (iii) stabilité *via* la conservation des normes et valeurs traditionnelles (« *conventionnalisme* »). Mais MSC et RWA ne sont pas isomorphiques : la MSC est un indice de motivations chroniques chez l'individu, le RWA est une mesure d'attitudes favorables aux normes, valeurs et règles traditionnelles ainsi qu'au système d'autorité qui les porte. En support, des études ont montré que la MSC n'est pas corrélée avec les préjugés, contrairement au RWA (Altemeyer, 1998 ; Duriez & Van Hiel, 2002 ; Heaven, Organ, Supavadeeprasit, & Leeson, 2006).

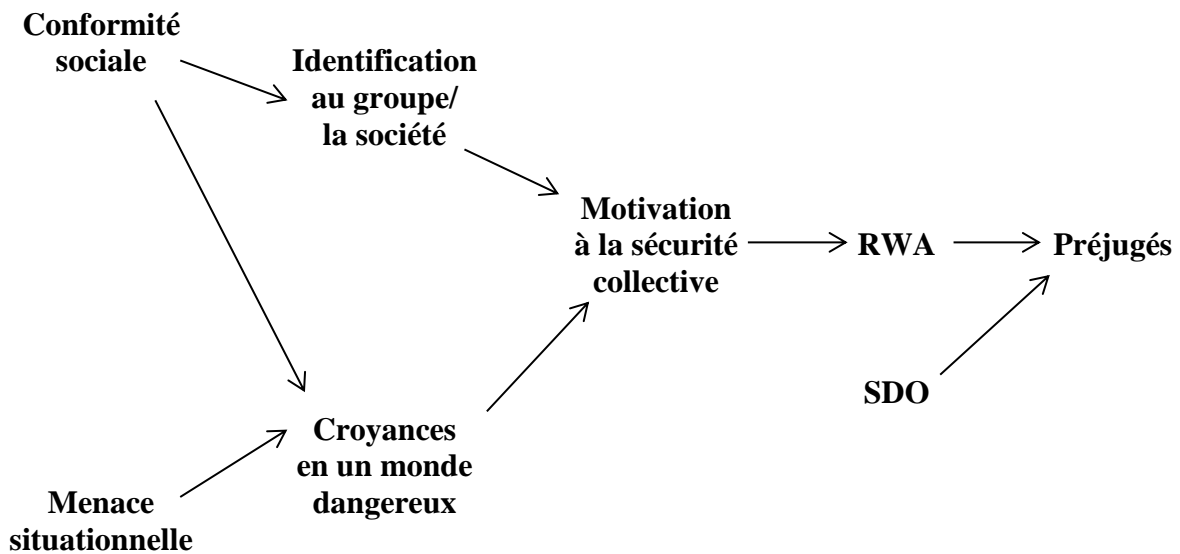


Figure 7. Modèle de la sécurité collective des attitudes autoritaires (Jugert & Duckitt, 2009). Le RWA serait une conséquence directe de la motivation à la sécurité collective, elle-même influencée par la personnalité (dont l'effet est médiatisé par l'identification groupale et la croyance en un monde dangereux) et la menace sociale (dont l'effet est médiatisé par la croyance en un monde dangereux). (Adapté de Jugert & Duckitt, 2009)

Recherche sur l'autoritarisme en contexte français

Les travaux réalisés sur l'autoritarisme en contexte français ont répliqué plusieurs résultats importants obtenus dans d'autres contextes culturels (e.g., Amérique du Nord, Nouvelle-Zélande, Angleterre, Allemagne, Italie). Par exemple, Gatto et Dambrun (2010) ont trouvé que (i) RWA et SDO prédisent les préjugés intergroupes, (ii) la manipulation expérimentale de la saillance de la menace augmente le niveau de RWA, (iii) le RWA médiatise l'effet de la menace sur les préjugés intergroupes. De manière cohérente avec l'approche motivationnelle développée par Duckitt et ses collaborateurs, Gatto et Dambrun (2010) ont proposé que les personnes placées dans un contexte menaçant éprouvent le besoin de restaurer un sentiment de sécurité favorisant de forts scores de RWA et l'expression de préjugés intergroupes. Les auteurs précisent que ce type de besoins apparaît tout

particulièrement dans des contextes valorisant la conformité sociale (e.g., institution policière). D'autres études ont montré l'influence du contexte social/groupal sur les attitudes autoritaires et les préjugés intergroupes : (i) l'incertitude sécuritaire favorise l'expression du RWA et des préjugés intergroupes, (ii) un contexte compétitif favorise l'expression du SDO et des préjugés intergroupes (e.g., Dru, 2007).

Dans leur ensemble, ces travaux suggèrent que le RWA serait l'expression dans la sphère sociale d'une plus grande vulnérabilité au stress et d'une préférence pour les stratégies défensives face aux défis environnementaux. Or, le paradigme de Milgram impose aux participants une situation de dilemme social propice à l'expérience du stress. Nous présenterons ci-dessous deux études portant sur l'influence des compétences socio-émotionnelles sur l'OD. De manière générale, les compétences socio-émotionnelles réfèrent aux capacités des individus à exprimer, réguler et comprendre les émotions éprouvées par soi et par autrui (e.g., Decety, 2010). Ces trois composantes ont suscité de nombreux travaux en psychologie expérimentale et en neurosciences au cours des deux dernières décennies et sont considérées comme des « acquisitions de bases » sur lesquelles l'empathie et le raisonnement moral vont se développer (e.g., Batson, 1991 ; Decety & Jackson., 2004).

1.1.3.2.3. Influence des compétences socio-émotionnelles sur l'OD

De récentes études attestent d'un intérêt émergent pour l'influence des compétences socio-émotionnelles sur l'OD. Ainsi, Burger (2009) s'est intéressé à l'influence de l'*empathie dispositionnelle* (Davis, 1994) en proposant l'existence d'une relation inverse entre propension des individus à l'empathie et propension à l'OD. L'auteur a réalisé une réplification « partielle » du paradigme de Milgram (variante 5, avec arrêt de l'expérience à 150 volts) et a mesuré l'empathie dispositionnelle des participants à l'aide de l'Indice de Réactivité Interpersonnelle (IRI ; Davis, 1983). Ce questionnaire distingue quatre dimensions de

l'empathie : (i) souci empathique (tendance à éprouver de la sympathie et de la compassion face à la détresse d'autrui), (ii) détresse personnelle (tendance à éprouver un sentiment d'inconfort face à la détresse d'autrui), (iii) prise de perspective (tendance à adopter la perspective d'autrui) et (iv) fantaisie (tendance à se placer soi-même dans des situations imaginaires). Les résultats ne montrent aucune relation entre les scores des participants aux différentes sous-échelles de l'IRI et l'OD. En revanche, ils suggèrent une relation inverse entre le score de souci empathique et la réticence à obéir (mesurée par le moment de formulation par l'expérimentateur de la première injonction) : les participants les plus « empathiques » semblent avoir exprimé plus tôt leur réticence à obéir.

Zeigler-Hill, Southard, Archer et Donohoe (2013) ont utilisé un paradigme d'obéissance impliquant l'administration de chocs sonores à la victime (de 60 à 150db, avec une incrémentation de 10db), et ont fait l'hypothèse que la réticence des individus à obéir dépendrait de l'intensité des *émotions négatives* éprouvées pendant l'expérience : plus les réactions émotionnelles aversives des participants seraient fortes, plus la propension à désobéir augmenterait². Pour tester cette hypothèse, les auteurs ont demandé aux participants de compléter un questionnaire portant sur leur état émotionnel à la fin de l'expérience (cf. tableau 6). Les auteurs se sont également intéressés à l'influence du *Neuroticisme* sur l'OD, trait de personnalité associé à une plus grande vulnérabilité au stress (e.g., Costa & McCrae, 1995), une plus forte propension à éviter l'incertitude, à adhérer aux normes (Allik & McCrae, 2004) et à se conformer à diverses formes de pression sociale (Drake, 2010). D'après Zeigler-Hill et al. (2013), les participants présentant un fort score de Neuroticisme devraient se montrer plus obéissants du fait d'une plus forte appréhension quant aux conséquences sociales du non-respect des demandes de l'autorité (e.g., désapprobation sociale). Les auteurs ont formulé une hypothèse d'interaction entre émotions négatives et Neuroticisme : les

² Les auteurs font référence au modèle de Schwarz et Clore (1996) voulant que les affects jouent un rôle important dans la manière dont les individus évaluent les situations nouvelles.

individus présentant un faible score de Neuroticisme et éprouvant le plus d'émotions négatives durant l'expérience devraient être les plus disposés à désobéir, du fait d'une moindre habitude à éprouver des émotions négatives et d'une évaluation moins négative des éventuelles conséquences sociales de leur désobéissance. Les résultats ne montrent aucune relation entre émotions négatives auto-reportées et OD, ou entre Neuroticisme et OD. En revanche, les auteurs ont trouvé (i) une relation significative et positive entre émotions négatives auto-reportées et réticence à obéir (mesurée par le nombre d'injonctions formulées par l'expérimentateur), (ii) une relation négative entre émotions négatives auto-reportées et Neuroticisme, (iii) un effet d'interaction entre émotions négatives auto-reportées et Neuroticisme suggérant que la relation entre émotions négatives auto-reportées et réticence à obéir n'existe que pour les participants ayant un faible score de Neuroticisme. Ainsi, il semble que les individus ayant exigé le plus grand nombre d'injonctions à poursuivre soient ceux présentant de faibles scores de Neuroticisme et des scores élevés d'émotions négatives auto-reportées.

Les résultats de ces deux études suggèrent (i) que les personnes ayant une plus forte tendance générale au souci empathique éprouvent une plus forte réticence à exécuter des ordres destructeurs (Burger, 2009), (ii) que les personnes présentant une moindre vulnérabilité au stress éprouvent plus d'émotions négatives durant l'expérience, celles-ci ayant pour conséquence une plus forte réticence à exécuter des ordres destructeurs (Zeigler-Hill et al., 2013). Ainsi la résonance empathique pourrait favoriser la désobéissance à des ordres destructeurs alors que la vulnérabilité au stress pourrait au contraire favoriser l'OD.

Cette revue de littérature sur les déterminants dispositionnels de l'OD a permis l'identification de plusieurs facteurs susceptibles d'influencer l'OD (autoritarisme de droite, empathie, vulnérabilité au stress). Elle met également en exergue un certain nombre de

problèmes chroniques dans la recherche sur les différences interindividuelles dans l'OD. Tout d'abord, ce champ souffre d'un *manque d'unité théorique et méthodologique* (i.e., les modèles théoriques mobilisés, les procédures expérimentales et la qualité des outils de mesure diffèrent selon les études ; e.g., Bègue et al., 2015 ; Blass, 1991). Il apparaît également que les hypothèses ne portent que très rarement sur des *mécanismes et processus*. Enfin, les quelques études ayant approché l'influence des émotions reposent sur l'usage de *mesures explicites* (e.g., Zeigler-Hill et al., 2013). Or, ce type de mesure est soumis à un certain nombre de biais et d'artefacts du fait (i) du contrôle par les participants de leurs réponses et (ii) de la faible accessibilité des participants à leurs processus et états internes (e.g., Blascovich, 2014). Nous terminerons cette partie par la présentation des principaux modèles théoriques de l'OD.

1.1.4. Modèles théoriques de l'OD

Hypothèse de l'état agentique

Selon Milgram (1974), toute théorie scientifique de l'obéissance doit s'efforcer de répondre à la question suivante : « Quelles sont les modifications qui se produisent quand l'individu autonome est inséré dans une structure sociale où il fonctionne non plus de façon indépendante, mais en tant que partie intégrante d'un système ? »

Milgram a articulé son raisonnement sur les causes de l'obéissance autour d'un nombre restreint de points : (i) la vie en groupe présente un avantage adaptatif en favorisant la survie des individus, (ii) les traits comportementaux et psychologiques produisant cette capacité à vivre en groupe sont la conséquence de l'histoire évolutive de l'espèce humaine, (iii) du point de vue de la cybernétique, la condition indispensable à l'insertion d'organismes indépendants dans une structure hiérarchique est la « suppression des directives du contrôle individuel en faveur de celui de l'agent coordinateur », (iv) les structures hiérarchiques ne peuvent fonctionner que si une modification est apportée aux éléments qui les composent, (v)

les individus s'intégrant dans ces structures hiérarchiques sont nécessairement modifiés dans leur mode de fonctionnement.

Pour Milgram (1974), ces modifications du fonctionnement des individus seraient la cause du comportement d'OD observé dans ses expériences : l'insertion dans une structure hiérarchique provoquerait la suspension du contrôle local en faveur des directives formulées par l'expérimentateur. Relativement à la question de la « localisation » de ce levier permettant le passage du mode autonome au mode systématique, Milgram (1974) propose la réponse suivante : « Tout comme chez les automates, il y a sûrement une modification de l'équilibre interne qui se réduit sans aucun doute à des *changements dans les schémas du fonctionnement neural*. Les inhibiteurs et les désinhibiteurs chimiques compromettent la possibilité d'action de certains mécanismes neuraux, mais dans quelle mesure exacte, nous sommes techniquement incapables de le déterminer au niveau chimio-neurologique. Nous avons cependant accès à une expression phénoménologique de ce délicat changement parce qu'il entraîne une modification d'attitude. » L'insertion d'un individu dans un système d'autorité induirait ainsi des altérations profondes dans l'équilibre interne et le comportement de cet individu, qui se percevrait alors non plus comme l'auteur de ses actions mais au contraire comme l'agent exécutif des volontés d'autrui. C'est ce que Milgram a appelé *état agentique*, défini comme « la condition de l'individu qui se considère comme l'agent exécutif d'une volonté étrangère », par opposition à l'*état autonome* dans lequel l'individu se perçoit comme l'auteur de ses actions. L'état agentique serait incompatible avec le sentiment de responsabilité mais également avec les considérations morales guidant le comportement des individus en état autonome.

Enfin, Milgram (1974) précise que si l'individu peut choisir de ne pas entrer en état agentique, la probabilité d'un tel refus est extrêmement ténue. Dans la perspective

milgramienne, l'obéissance est en effet à la fois innée et fortement dépendante des influences de l'environnement : l'évolution aurait doté l'humain de certaines structures « mentales » permettant l'obéissance, et l'expérience (*i.e.*, apprentissage culturel) serait responsable de l'apprentissage de la docilité. Le cumul d'une prédisposition à l'obéissance et de la socialisation au sein de groupes hiérarchiques serait responsable de la faible probabilité d'un maintien d'un état autonome après insertion dans un système d'autorité.

Si l'hypothèse de l'état agentique avancée par Milgram offre des pistes importantes de compréhension de l'OD, elle ne permet cependant pas d'expliquer les différences interindividuelles dans ses expériences : les participants étaient tous issus de la même culture et aucune influence n'a pu être attribuée aux variables sociodémographiques.

Désengagement moral

Bandura (1986, 1991, 1999) a développé une théorie du « désengagement moral » proposant que le comportement d'atteinte à autrui serait la conséquence d'un désengagement des mécanismes d'auto-sanction morale. Cette théorie s'inscrit dans le cadre d'une approche « sociocognitive » du fonctionnement moral reposant sur la notion d'agentivité, définie comme la capacité humaine à influencer intentionnellement sur le cours de sa vie et de ses actions. L'agentivité morale posséderait une forme inhibitive et une forme proactive : (i) l'agentivité « inhibitive » permettrait d'empêcher le recours à un comportement destructeur, (ii) l'agentivité « proactive » s'exprimerait dans le comportement prosocial.

L'auteur propose que le développement moral se caractériserait par le passage de l'hétéro-régulation à l'autorégulation : (i) durant les premières années de vie, le comportement moral serait régulé par autrui et soumis à des sanctions sociales, (ii) au cours de sa socialisation, l'individu internaliserait des standards moraux et mettrait en place des mécanismes d'auto-sanction. Les sanctions positives viendraient renforcer les comportements

ajustés aux standards moraux internalisés, alors que les sanctions négatives seraient aversives et viendraient inhiber le recours à des comportements transgressant ces standards moraux. Ce mécanisme d'autorégulation posséderait une dimension anticipatrice : dans une situation exerçant une pression à l'action destructrice, l'auto-influence permettrait aux individus de ne pas se conformer. L'exercice continu de cette auto-influence motiverait et régulerait le comportement moral. Cependant, les standards moraux n'interviennent pas toujours comme régulateurs internes du comportement. Les mécanismes d'autorégulation ne pourraient intervenir qu'à condition d'être « activés », et il existerait un certain nombre de processus psychosociaux responsables d'un désengagement du contrôle personnel. Activation sélective et désengagement du contrôle personnel seraient à l'origine de l'adoption de comportements destructeurs.

Bandura (1986, 1999) a proposé que l'auto-sanction peut être désengagée à différents moments du processus d'autorégulation : (i) reconstruction de la *perception du comportement d'atteinte à autrui* afin de lui conférer une valeur morale, (ii) minimisation par l'individu de son *rôle dans l'action d'atteinte à autrui*, (iii) minimisation des *conséquences* du comportement d'atteinte, (iv) reconstruction de la *perception de la victime* de manière à lui attribuer la responsabilité de l'atteinte subie (*cf.* figure 8). Le premier niveau implique des mécanismes cognitifs relatifs à la justification morale de l'atteinte à autrui (e.g., protéger la sécurité de l'État, faire avancer la Science ; Kelman, 1973 ; 2005), à l'utilisation d'un vocabulaire euphémisant (e.g., opération de nettoyage, mission de pacification ; Gambino, 1973), et au recours à des comparaisons induisant un contraste favorable à l'action d'atteinte (e.g., utilitarisme moral ; Bandura, 1999). Le deuxième niveau met en jeu des mécanismes cognitifs permettant le déplacement ou la diffusion de la responsabilité individuelle (e.g., état agentique ; Milgram, 1974). Cette proposition s'appuie en partie sur les résultats de Milgram (1974) relatifs à l'influence de la proximité de l'expérimentateur. Le troisième niveau

implique des mécanismes cognitifs permettant l'évitement d'une confrontation directe aux conséquences pour autrui du comportement d'atteinte, ou la réévaluation de ces conséquences dans le sens d'une sous-estimation. Cette proposition s'appuie pour partie sur les résultats de Milgram relatifs à l'influence de la proximité de la victime. Enfin, le quatrième niveau met en jeu des processus cognitifs permettant d'attribuer à la victime (et non à soi) la responsabilité de l'atteinte subie. Ce blâme porté à la victime peut prendre des formes extrêmes comme la déshumanisation : la victime se voyant dénier sa pleine appartenance à l'espèce humaine, elle sera exclue du domaine des obligations morales (Bandura, 1999 ; Kelman, 2005 ; Opatow, 1990).

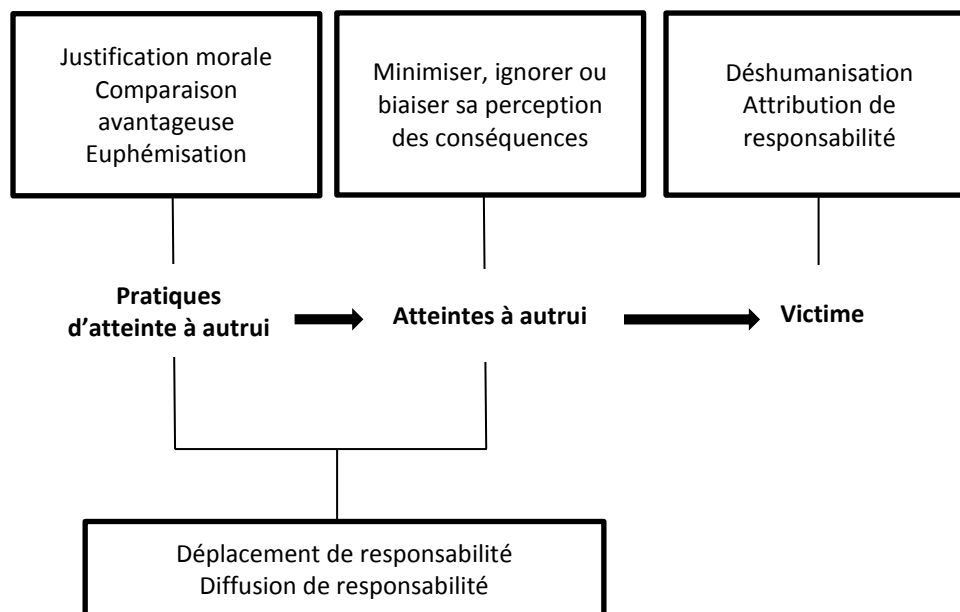


Figure 8. Mécanismes par lesquels l'auto-sanction morale est sélectivement engagée ou désengagée à différents points du processus d'autorégulation. (Bandura, 1986)

Les résultats d'un récent travail de recherche (Caspar, Christensen, Cleeremans, & Haggard, 2016) apportent un soutien empirique à l'idée que l'insertion dans un contexte coercitif altère le sentiment d'agentivité (*i.e.*, expérience subjective de contrôler ses actions ; Bandura, 1999 ; Milgram, 1974). Dans une première étude, Caspar et al. (2016) se sont

intéressés à l'influence de la coercition sur une mesure implicite du sentiment d'agentivité : la perception de l'intervalle temporel entre une action et sa conséquence (cet intervalle est perçu comme plus court en cas d'action intentionnelle vs mouvement passif ; e.g., Haggard, Clark, & Kalogeras, 2002). Les auteurs ont utilisé un paradigme d'obéissance impliquant l'administration d'un choc électrique modéré à autrui dans le cadre d'un jeu économique. Dans une première condition, les participants devaient suivre les commandes de l'expérimentateur (condition « coercition »). Dans une seconde condition, les participants décidaient eux-mêmes d'administrer ou non un choc (condition « libre choix »). Les résultats montrent une estimation plus importante de l'intervalle action-conséquence dans la condition « coercition », suggérant qu'agir sur la commande d'autrui réduit l'expérience subjective de l'agentivité. Dans une seconde étude, Caspar et al. (2016) ont utilisé l'électroencéphalographie pour mesurer l'influence de la coercition sur les réponses neurales des participants aux conséquences de leurs actions. Les résultats montrent une augmentation d'amplitude plus importante du potentiel évoqué auditif N1 pour les conséquences des actions dans la condition « libre choix », suggérant que la coercition diminue le traitement des conséquences de l'action.

L'OD, un comportement « négocié » ?

L'approche socio-historique de l'OD appréhende l'usage de l'autorité comme une forme « légitime » d'influence (e.g., Kelman, 2001, 2005 ; Kelman & Hamilton, 1989). Selon ces auteurs, les individus ne seraient pas passifs face à l'usage de l'autorité mais procéderaient au contraire à une évaluation active de sa légitimité à trois niveaux : (i) au niveau du système d'autorité au sein duquel s'établit la relation entre les individus et une figure d'autorité, (ii) au niveau de la figure d'autorité elle-même (e.g., autorité bureaucratique, scientifique), (iii) au niveau de la demande formulée par l'autorité. L'obéissance se produirait seulement si la

légitimité est satisfaite à ces trois niveaux. Dans cette perspective théorique, l'obéissance serait un comportement volontaire et non la conséquence d'une conformation passive dans le cadre d'une relation coercitive (voir aussi Passini & Morselli, 2009).

Burger, Girgis et Manning (2011) se sont intéressés aux réactions des participants aux quatre injonctions formulées par l'expérimentateur dans les expériences de Milgram. Selon ces auteurs, seule la quatrième injonction constitue un ordre direct (« Vous n'avez pas d'autres choix, vous devez continuer »). Par contraste, les trois premières injonctions représenteraient plutôt des « arguments » quant au pourquoi de la nécessité de poursuivre l'expérience. Les auteurs notent que l'expérimentateur n'utilise aucun terme explicitement relié à l'obéissance (« ordre », « commande », ou « obéir »). Burger et al. (2011) ont calculé le taux d'obéissance à chaque injonction et ont trouvé que « plus la déclaration de l'expérimentateur ressemble à un ordre, moins les participants tendent à obéir ». De plus, il semble qu'aucun participant n'ait obéi à la quatrième injonction. Selon les auteurs, ces résultats suggèrent que les participants ne suivent l'expérimentateur que si celui-ci justifie son action (en termes de bénéfices scientifiques) et s'opposent à l'idée voulant que les participants auraient « obéi à des ordres ». Selon Reicher et Haslam (2011), l'utilisation d'injonctions autoritaires serait contre-productive car elle saperait le sentiment d'identité partagée et donc la motivation à coopérer.

Hypothèse de l'identification sociale

Reicher et Haslam (2011) ont proposé que l'obéissance repose sur la propension des individus à s'identifier à l'autorité : l'identification à l'expérimentateur (et plus largement à la communauté scientifique) devrait augmenter l'obéissance, alors que l'identification à la victime (et à la communauté générale) devrait favoriser la désobéissance. Selon les auteurs, la manipulation de facteurs « spatiaux » dans les expériences de Milgram (*i.e.*, proximité de

l'expérimentateur et de la victime) aurait pu impacter la formation d'alliances ou de « coalitions » (participant-expérimentateur *vs* participant-élève) et ainsi déterminer le taux d'obéissance. Pour tester l'hypothèse d'une relation entre identification sociale et obéissance, Reicher, Haslam et Smith (2012) ont réalisé une étude consistant à soumettre aux participants une description de la condition standard de Milgram et de 12 variantes, et de leur demander d'évaluer dans quelle mesure chacune des variantes pouvait conduire à s'identifier à l'expérimentateur (et plus largement à la communauté scientifique) et à l'élève (et à la communauté générale). Les résultats montrent une corrélation positive entre identification à l'expérimentateur et les taux d'obéissance obtenus dans chaque variante, et une corrélation négative entre identification à l'élève et taux d'obéissance. Pour les auteurs, ces résultats remettent en cause l'hypothèse de l'état agentique en suggérant que l'OD dans le paradigme de Milgram serait non pas la conséquence d'une « simple obéissance » mais le résultat d'une identification active à l'expérimentateur et sa mission. Si cette approche peut contribuer à une meilleure compréhension des déterminants de l'OD, il est important de relever le fait que l'hypothèse formulée par Reicher et Haslam (2011) n'a jamais été vérifiée dans le cadre d'un paradigme d'obéissance.

Si les modèles théoriques actuels de l'OD peuvent contribuer à une meilleure compréhension de l'OD, ils présentent un certain nombre de faiblesses. Tout d'abord, ces modèles n'intègrent pas les facteurs dispositionnels. Plus fondamentalement, il apparaît que les auteurs aient privilégié la cognition et la rationalité au détriment des émotions comme guide du comportement. Ce choix s'ancre dans un contexte académique et extra-académique de sous-estimation du rôle des émotions dans la détermination du comportement humain. Comme l'a souligné de Waal (2003), l'étude de la moralité humaine a vu s'opposer deux courants, l'un arguant d'une discontinuité entre espèce humaine et animaux sociaux (approche discontinuiste mettant l'accent sur la rationalité), et l'autre insistant sur une continuité

évolutive au regard des nombreuses similitudes existant entre humain et primates non-humains (approche continuiste accordant une place centrale aux émotions sociales). L'approche discontinuiste est très influente dans l'étude de la moralité. Elle s'inscrit dans une certaine tradition philosophique (e.g., Hobbes, 1651) et a été renforcée par des biologistes comme Huxley (1894). Nous présenterons ci-dessous (de manière synthétique) les approches discontinuiste et continuiste dans l'étude de la moralité humaine.

1.1.5. Approches discontinuiste et continuiste de la moralité humaine

Approche discontinuiste

L'idée voulant que « l'Homme est un loup pour l'Homme » (*Homo homini lupus* ; Hobbes, 1651) a été portée par un courant très influent de la biologie initié par Huxley (1894), dont la principale proposition est que l'humain serait fondamentalement « malveillant et égoïste ». Dans cette perspective, l'humain ne peut accéder à une quelconque forme de moralité qu'au prix d'efforts importants. Selon la théorie du contrat social développée dans un premier temps par Hobbes (1651), la moralité et la socialité humaines sont la conséquence tardive de processus conscients, contrôlés et coûteux aboutissant à des choix raisonnés. Dans la continuité de cette approche philosophique, Huxley (1894) a conceptualisé la moralité humaine comme une victoire culturelle relativement récente sur les processus évolutifs. Plus récemment, Williams (1988) a soutenu que la moralité humaine serait un sous-produit du processus évolutif, une capacité « accidentelle » générée par un processus biologique normalement opposé à l'expression d'une telle capacité. La moralité serait ainsi un vernis fragile apposé sur les passions humaines inévitablement antisociales, amORAles et égoïstes. Selon Wright (1994), l'humain est potentiellement mais non naturellement moral.

Approche continuiste

Les travaux sur les origines évolutives du comportement social suggèrent au contraire que la morale serait la conséquence d'instincts sociaux communs aux primates sociaux. Selon des auteurs comme de Waal (1996, 2003, 2008), la pression sélective a conduit au développement des *émotions sociales* (émotions induites dans un contexte d'interaction avec un ou plusieurs individus) dans le contexte des soins parentaux, et ce bien avant l'apparition de l'espèce humaine. Savoir protéger et prendre soin de la progéniture représente un défi adaptatif majeur pour la survie d'une espèce, et ce défi est exacerbé chez l'humain dont le nouveau-né est particulièrement vulnérable et dépendant. Dans ce contexte, la capacité à faire l'expérience et comprendre les états émotionnels et corporels du nourrisson représente un avantage adaptatif évident en permettant de réagir de manière rapide et adapté aux signaux de détresse. Du fait de sa valeur adaptative, cette capacité générale d'*empathie* se serait ensuite élargie à l'ensemble des interactions entre individus, constituant une base émotionnelle pour le développement de la morale. Si la moralité constitue une addition tardive au comportement humain, elle devrait être supportée par les aires corticales les plus récentes. Or, les travaux réalisés en neurosciences de la morale montrent que le jugement moral mobilise des processus émotionnels automatiques sous-tendus par des structures cérébrales anciennes du point de vue évolutif (e.g., Decety & Svetlova, 2012 ; Moll et al., 2005 ; Preston & de Waal, 2002).

Nous présenterons dans le deuxième chapitre de cette introduction les apports des neurosciences à la compréhension du fonctionnement moral dans des situations impliquant une atteinte à autrui. Nous aborderons dans un premier temps les neurosciences de la morale (de manière synthétique) puis les neurosciences sociales de l'empathie (de manière plus extensive). Nous mettrons l'accent sur le rôle respectif des mécanismes de résonance et d'autorégulation, du système nerveux central (SNC) et du système nerveux autonome (SNA).

Ce chapitre sera l'occasion d'une présentation de l'étude IRMf de Cheetham et al. (2009) portant sur la nature de l'expérience émotionnelle des participants au cours de la procédure d'obéissance.

CHAPITRE 2. APPORTS DES NEUROSCIENCES SOCIALES

1.2.1. Neurosciences de la morale

1.2.1.1. Approche dualiste de la morale

Les individus sont régulièrement confrontés à des situations impliquant un conflit entre deux traditions de l'éthique normative : (i) l'*utilitarisme* soutient qu'une action est juste si elle augmente le bien-être du plus grand nombre, sans considération pour la sévérité de cette action, (ii) le *déontologisme* affirme qu'une action est juste si elle est conforme à certains impératifs éthiques (e.g., ne pas porter atteinte à autrui ; Crockett, 2013). Les dilemmes moraux utilisés dans la recherche sur le jugement moral se caractérisent par la présence d'un conflit moral dont l'issue pourrait se traduire par une atteinte sévère à autrui. Le « dilemme du tramway » est l'un des plus utilisés. Ce dilemme met en scène un tramway dévalant une voie ferrée et se dirigeant vers cinq personnes. Dans la version « aiguillage » (dilemme moral impersonnel), le seul moyen de les sauver est d'actionner un aiguillage qui orientera le tramway sur une voie où ne se trouve qu'une seule personne, qui de ce fait perdra la vie. Dans la version « pont » (dilemme moral personnel), le seul moyen de sauver les cinq personnes est de pousser une autre personne du haut d'un pont afin de la précipiter sur les rails. Son corps permettra de stopper le tramway, épargnant ainsi les cinq personnes. Alors que la majorité des participants favorise l'option utilitariste dans la version « aiguillage », seule une minorité présente une préférence pour l'option utilitariste dans la version « pont » (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001).

1.2.1.2. Substrats neurobiologiques

Des études en imagerie cérébrale (e.g., Greene et al., 2001) ont montré que les dilemmes moraux personnels provoquent une activité hémodynamique plus importante au

sein des circuits neuronaux impliqués dans les réactions viscérales/émotionnelles automatiques (e.g., amygdale : AMY ; cortex préfrontal ventromédian : CPFVM ; cortex orbitofrontal : COF), alors que les dilemmes moraux impersonnels induisent un gain d'activité au sein des circuits neuronaux en charge des processus cognitifs/rationnels contrôlés (e.g., cortex préfrontal dorsolatéral : CPFDL ; cf. figure 9). D'autres études ont montré que la compromission des différentes composantes du « cerveau moral affectif » (AMY, CPFVM et COF) augmente l'utilitarisme moral (Blair, 2007 ; Funk & Gazzaniga, 2009 ; Moll & de Oliveira-Souza, 2007).

Ces résultats sont utilisés pour soutenir l'approche dualiste de la morale proposant que les processus contrôlés rationnels sous-tendent les décisions utilitaristes (e.g., Paxton, Ungar, & Greene, 2012) alors que les réactions émotionnelles automatiques favoriseraient le déontologisme (e.g., Greene, 2009). L'équilibre entre les différentes structures cérébrales sollicitées lors de la résolution de dilemmes moraux conditionnerait la prise de décision: (i) une augmentation de l'influence inhibitrice du CPFDL sur le « cerveau moral affectif » favoriserait les décisions utilitaristes, (ii) lorsque les circuits neuronaux impliqués dans les réactions émotionnelles automatiques supplantent les circuits en charge des processus rationnels contrôlés, les options déontologiques sont privilégiées. La recherche récente a cependant remis en cause le modèle dualiste en suggérant qu'un déficit de souci empathique serait un prédicteur plus robuste de l'utilitarisme que l'intégrité des processus cognitifs/rationnels. De récentes études ont mis en évidence que l'utilitarisme est associé au souci égoïste (e.g., Kahane, Everett, Earp, Farias, & Savulescu, 2015), à des traits de personnalité antisociaux comme la psychopathie et le machiavélisme (e.g., Bartels & Pizarro, 2011), et à une moindre aversion pour l'atteinte à autrui (e.g., Crockett et al., 2010). Des études lésionnelles ont montré qu'une atteinte au niveau CPFVM provoque un biais utilitariste (e.g., Koenigs et al., 2007). Moll et ses collaborateurs ont proposé que cette aire

cérébrale participe à la génération des émotions prosociales (e.g., culpabilité, compassion) et que les personnes présentant des lésions du CPFVM accepteraient de sacrifier un individu pour en sauver d'autres du fait de leur moindre capacité à éprouver des émotions prosociales (e.g., Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005).

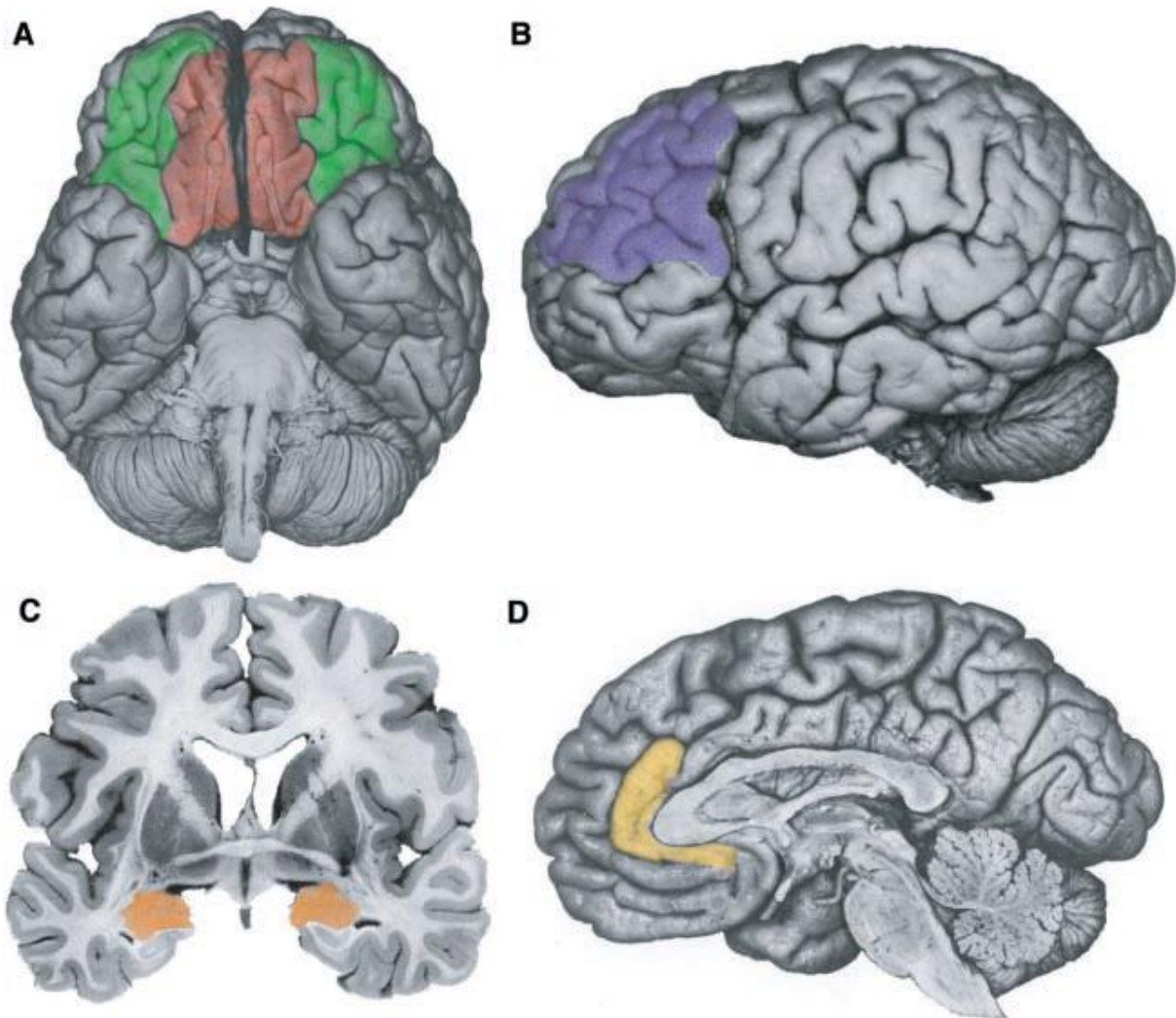


Figure 9. Structures cérébrales impliquées dans le jugement moral. (A) Le cortex orbitofrontal apparaît en rouge et le cortex préfrontal ventromédian en vert. (B) Cortex préfrontal dorsolatéral. (C) Amygdale. (D) Cortex cingulaire antérieur. (Selon Davidson, Putnam, & Larson, 2000)

Amygdale cérébrale

L'AMY est en charge de la production des réponses émotionnelles automatiques aux stimuli aversifs. Elle est impliquée dans le conditionnement aversif et dans la socialisation morale (e.g., Moll et al., 2005). Une moindre activité au sein du système cortico-limbique incluant l'AMY a été associée à une moindre aversion pour l'atteinte à autrui (e.g., Moll & de Oliveira-Souza, 2007). Des travaux ont montré que la psychopathie est associée à une moindre activité amygdalienne dans les situations mettant en jeu l'intégrité physique d'autrui, et à une difficulté à l'apprentissage moral du fait d'une incapacité à associer l'atteinte à autrui à des réponses émotionnelles aversives (e.g., Blair, 2007).

Cortex préfrontal ventromédian et cortex orbitofrontal

Le CPFVM est fortement connecté avec l'AMY, et possède des projections vers les régions basales et le tronc cérébral influençant les réactions viscérales aux stimuli émotionnels (Koenigs et al., 2007). Le COF est lui aussi fortement connecté avec d'autres régions cérébrales : (i) connexion avec le cortex perceptif permettant l'évaluation de l'environnement et des conséquences de l'action, (ii) connexion avec le cortex frontal dorsal permettant la planification de l'action, (iii) connexion avec l'insula permettant l'accès aux sensations viscérales, (iv) connexion avec l'AMY permettant d'anticiper les conséquences émotionnelles des actions. Le CPFVM et le COF sont des composantes importantes d'un réseau cortico-limbique permettant aux individus d'associer expérience émotionnelle et évaluation/décision morale (e.g., Moll et al., 2005). De nombreuses études montrent que la compromission ou l'inhibition de ces aires cérébrales (notamment au niveau de l'hémisphère droit) sont à l'origine d'une large gamme de déficits socio-émotionnels (e.g., Eslinger, 1998 ; Koenigs et al., 2007 ; Shamay-Tsoory et al., 2004). Ainsi, des études lésionnelles ont montré qu'une atteinte au niveau du COF provoque un déficit sévère de la capacité à percevoir,

partager et comprendre les émotions d'autrui, et il a été proposé que le COF est en charge de la composante affective de la « théorie de l'esprit » (*i.e.*, capacité à attribuer à autrui des émotions et sentiments ; Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). La compromission du COF a également été associée à un défaut de réponse émotionnelle face à des situations de transgression morale et d'atteinte à autrui (e.g., Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007 ; Namiki et al., 2008 ; Young, Bechara, Tranel, Damasio, Hauser, & Damasio, 2010). Dans leur ensemble, les travaux en neurosciences de la morale montrent que le CPFVM et le COF sont impliqués dans l'intégration des signaux socio-émotionnels à la prise de décision, ainsi que dans l'anticipation et l'évaluation des conséquences (corporelles et affectives) d'une décision morale.

Cortex préfrontal dorsolatéral

Le CPFDL joue un rôle important dans les processus cognitifs/rationnels contrôlés (e.g. Greene et al., 2004). Il est fortement impliqué dans le contrôle cognitif, le raisonnement, la mise en œuvre de principes moraux abstraits (e.g., Mansouri, Tanaka, & Buckley, 2009). Le CPFDL est plus généralement impliqué dans les fonctions exécutives et l'autorégulation (e.g., de Vignemont & Singer, 2006).

Cortex cingulaire antérieur

Le cortex cingulaire antérieur (CCA) est quant à lui activé lorsque les dilemmes moraux présentent une certaine difficulté (Funk & Gazzaniga, 2009). Cette région participe à la fois à la régulation émotionnelle et au contrôle cognitif. Le CCA est divisé en deux composantes : (i) cognitive (partie dorsale fortement connectée avec le cortex préfrontal – CPF) et (ii) émotionnelle (partie ventrale fortement connectée avec l'AMY). Cette région cérébrale joue un rôle central dans l'empathie pour la douleur (e.g., Decety & Lamm, 2009).

De nombreux auteurs ont proposé que les mécanismes de résonance empathique seraient à l'origine de l'aversion pour l'atteinte à autrui et du raisonnement moral orienté vers le bien-être d'autrui (e.g., Batson, 1991 ; Blair, 2007 ; Decety & Jackson, 2004 ; Preston & de Waal, 2002 ; Singer et al., 2004). Nous présenterons ci-dessous les principaux travaux réalisés en neurosciences sociales de l'empathie suggérant un rôle important des mécanismes de résonance et d'autorégulation dans l'initiation d'un comportement prosocial face à autrui en détresse. Nous présenterons ensuite l'étude IRMf de Cheetham et al. (2009) abordant l'OD dans la perspective des neurosciences sociales de l'empathie.

1.2.2. Neurosciences sociales de l'empathie

Résonance neurale et affective face à la douleur d'autrui

À un niveau phénoménologique, l'empathie renvoie à un *sentiment de similitude* entre les émotions exprimées par autrui et celles que l'on éprouve (e.g., Preston & de Waal, 2002). La recherche portant sur les réactions neurales à la douleur d'autrui a démontré que la simple perception de la douleur d'autrui provoque chez l'observateur l'activation de réseaux de neurones impliqués dans l'expérience personnelle de la douleur (« neuromatrice de la douleur » ; cf. figures 10 et 11). Par exemple, dans plusieurs études IRMf, les participants ont été exposés (i) à des vidéo clips montrant des individus exprimant de la douleur suite à une stimulation douloureuse (e.g., piqûre sur le bout d'un doigt), (ii) à la même stimulation douloureuse. Les résultats montrent une activation du CCA, de l'insula et du cortex somatosensoriel dans les deux conditions (e.g., Lamm et al., 2011). Un nombre important d'études a répliqué ces résultats, suggérant que l'observation et l'expérience de la douleur activent les composantes sensori-discriminantes et affectives-motivationnelles de la neuromatrice de la douleur (e.g., Jackson, Brunet, Meltzoff, & Decety, 2006 ; Jackson, Rainville, & Decety, 2006 ; Lamm, Nusbaum, Meltzoff, & Decety, 2007).

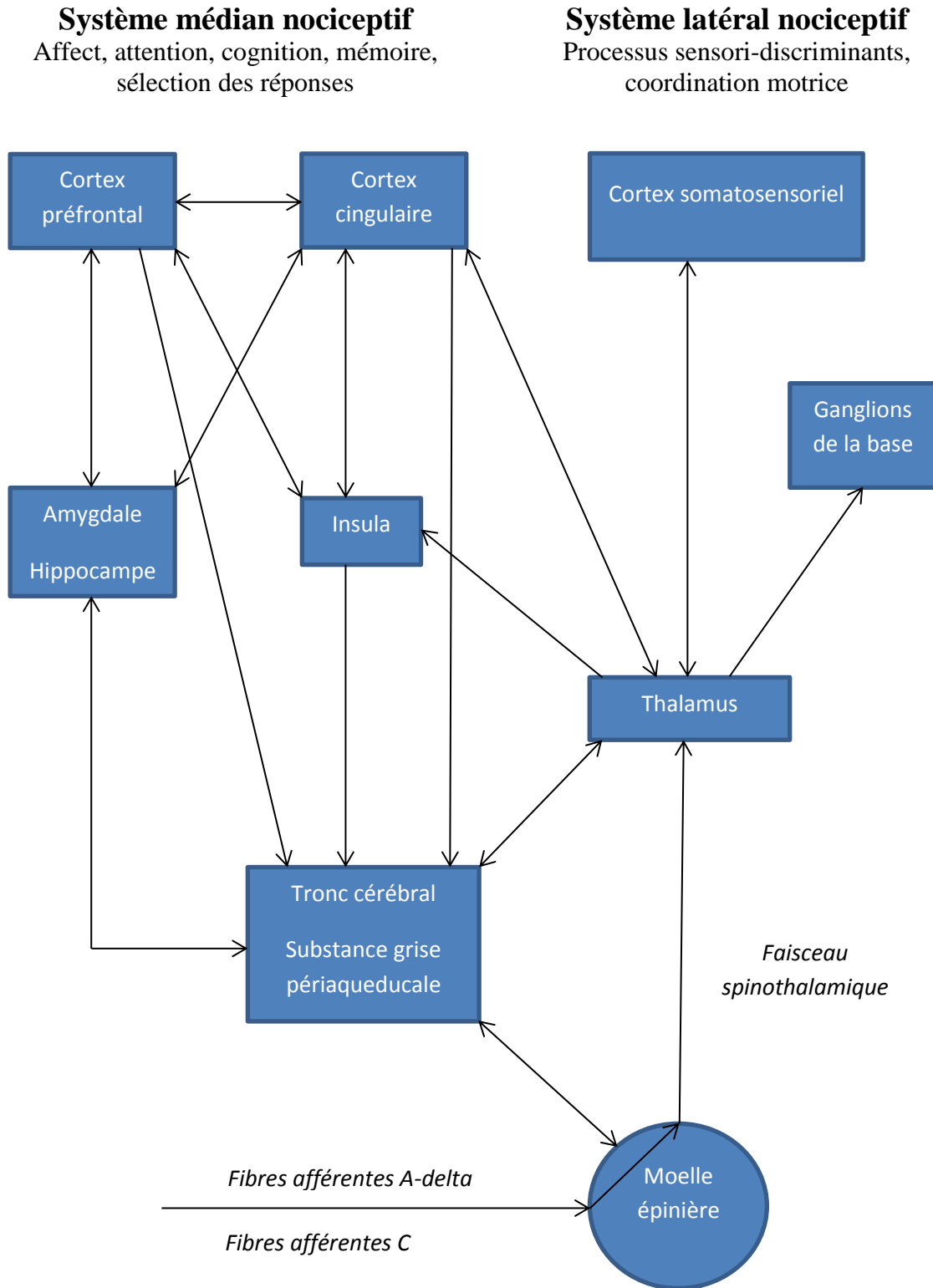


Figure 10. Représentation schématique des principales composantes anatomiques de la neuromatrice de la douleur. (Decety & Lamm, 2009)

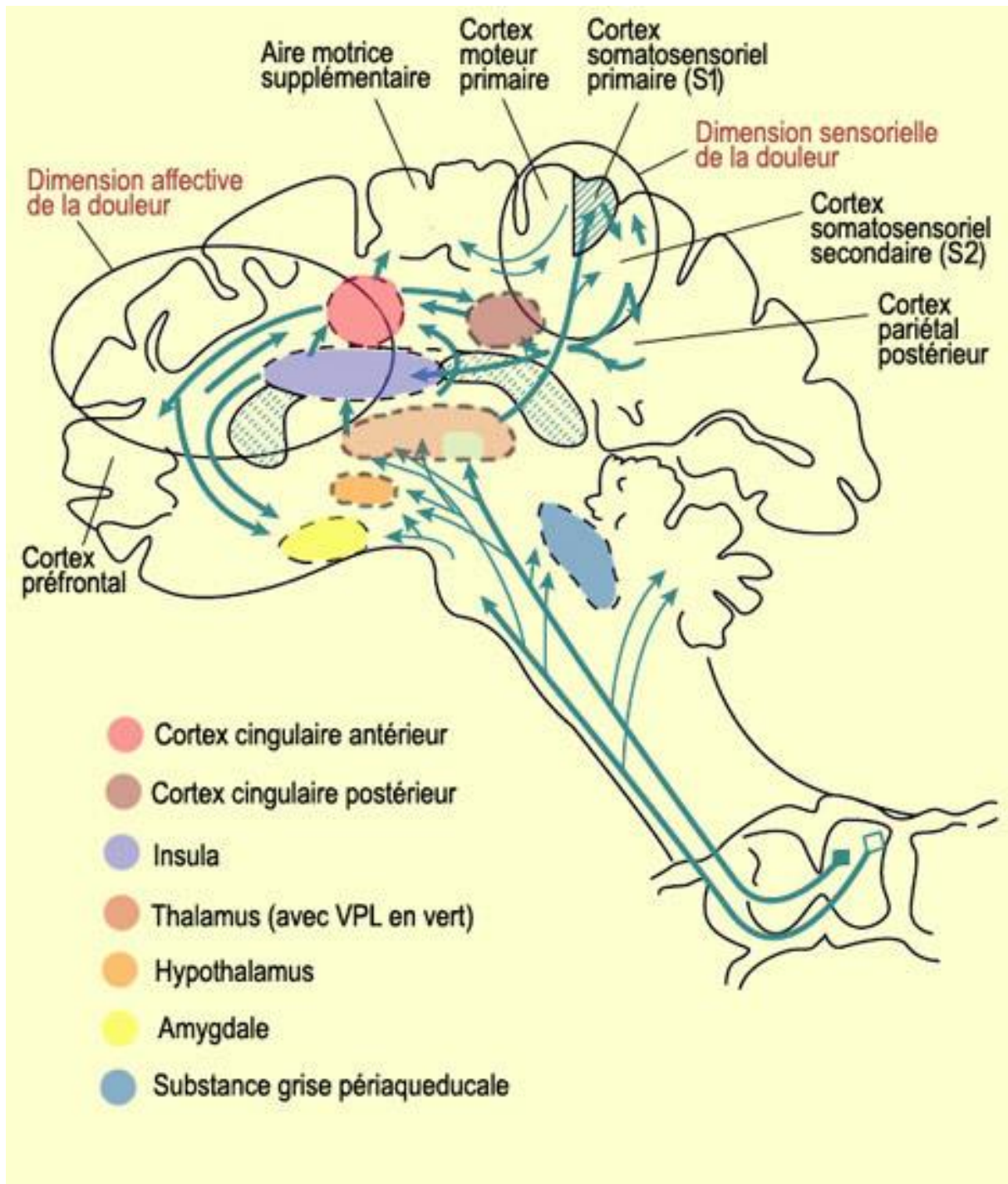


Figure 11. Distinction des composantes sensoridiscriminantes et affectives-motivationnelles de la neuromatrice de la douleur. Les cortex somatosensoriels primaire et secondaire sont impliqués dans les aspects sensori-discriminants de la douleur (localisation corporelle et intensité du stimulus). Le cortex cingulaire antérieur et l'insula sont impliqués dans la composante affective-motivationnelle de la douleur (évaluation de l'inconfort subjectif et préparation de la réponse dans un contexte aversif ou douloureux). (Decety & Lamm, 2009)

Cependant, les activations au sein de la neuromatrice de la douleur ne sont pas spécifiques à l'expérience émotionnelle de la douleur. En effet, de telles activations sont également impliquées dans l'évaluation de stimuli négatifs, l'attention portée à des stimuli nocifs ou encore la sélection de mouvements musculo-squelettiques défensifs (e.g., AMY, insula ; Decety & Lamm, 2009). Cette implication de la neuromatrice de la douleur dans l'évaluation de la menace et la sélection de réactions défensives (comme la fuite) peut aider à comprendre pourquoi l'observation de la douleur d'autrui ne génère pas automatiquement un comportement orienté vers autrui (e.g., aide).

Sympathie vs détresse personnelle

La recherche sur les conséquences affectives-motivationnelles du partage de la douleur a montré que l'observation de la douleur d'autrui ne provoque pas systématiquement des sentiments prosociaux. En effet, le partage de la douleur peut susciter soit de la *sympathie* (une réponse affective-motivationnelle orientée vers autrui) soit de la *détresse personnelle* (une réponse émotionnelle aversive centrée sur soi ; Batson, 1991 ; Eisenberg & Eggum, 2009 ; Decety & Lamm, 2009). Alors que la sympathie est considérée comme un précurseur du comportement prosocial (e.g., Eisenberg et al., 1996), la détresse personnelle a été associée à des sentiments de vulnérabilité, d'anxiété et la motivation à réduire non pas la détresse d'autrui mais son propre inconfort (Batson, 1991 ; Eisenberg et al., 2006). Les réactions affectives-motivationnelles face à la douleur d'autrui pourraient dépendre de la perspective adoptée par l'observateur. Des études IRMf (Jackson, Rainville, & Decety, 2006 ; Lamm, Batson, & Decety, 2007) ont montré des patterns d'activation neurale spécifiques suivant la perspective que l'on adopte face à la souffrance d'autrui. Ainsi, *s'imaginer être dans la même situation qu'autrui* provoque une élévation des sentiments de détresse personnelle associée à une forte augmentation de l'activité hémodynamique au niveau des structures cérébrales

codant les aspects affectifs-motivationnels de la douleur, comme l'AMY et la substance grise périaqueducale (structure impliquée dans la modulation de la douleur et la sélection de comportements défensifs). En revanche, *adopter la perspective d'autrui* induit une diminution des sentiments de détresse couplée à une baisse de l'activité hémodynamique au sein de l'AMY corrélée à une augmentation de l'activité du CPFDL, CPFVM et de la jonction temporo-pariétale droite. Ces résultats montrant que la prise de perspective s'accompagne d'activations additionnelles sont importants car ces trois aires corticales participent à la conscience de soi et à la distinction soi/autrui (Decety & Jackson, 2004).

Ces résultats sont primordiaux pour la compréhension des déterminants du comportement prosocial, et, de manière indirecte, des comportements liés à l'OD. Des auteurs ont proposé que la détresse personnelle soit la conséquence d'une confusion soi/autrui et que la sympathie résulte de l'intervention de mécanismes d'autorégulation maintenant une différenciation des perspectives. L'empathie semble ainsi impliquer à la fois des mécanismes rapides et automatiques de contagion émotionnelle et des mécanismes de plus haut niveau (contrôle cognitif/autorégulation) qui détermineront les conséquences affectives-motivationnelles et comportementales du partage affectif initial, de nature aversive dans le cas de la douleur.

Modèle de Decety et Jackson (2004)

L'empathie n'apparaît plus aujourd'hui comme un concept unitaire, référant à un mécanisme cognitif ou affectif circonscrit, et sous-tendu par un réseau cérébral spécifique. L'empathie est au contraire appréhendée comme un phénomène reposant sur divers processus affectifs et cognitifs sollicitant de nombreuses structures corticales et sous-corticales. Nous avons précisé au début de cette partie que l'empathie renvoyait à un sentiment de similitude entre les émotions que l'on éprouve et celles exprimées par autrui (e.g., Preston & de Waal,

2002). Decety et Jackson (2004) ont suggéré que l'empathie désigne un sentiment de similitude entre les émotions que l'on éprouve et celles exprimées par autrui *sans perte de distinction entre soi et autrui*. Ces auteurs ont défini un modèle proposant que l'empathie impliquerait trois compétences socio-émotionnelles : (i) le partage affectif, (ii) la conscience de soi, (iii) la flexibilité mentale et l'autorégulation. L'autorégulation est considérée comme une composante centrale en permettant la gestion et l'optimisation du partage de la douleur. Sans une intervention efficace des mécanismes d'autorégulation, les réponses somatiques et autonomes associées à la résonance affective provoqueront l'expérience de la détresse personnelle. Le tableau 7 présente les aspects fonctionnels et anatomiques de ces trois compétences.

Tableau 7

Compétences socio-émotionnelles associées à l'empathie selon Decety et Jackson (2004)

Compétences socio-émotionnelles	Aspects fonctionnels	Aspects anatomiques
Partage affectif (ou contagion émotionnelle)	Réfère au phénomène de résonance affective résultant de l'activation automatique de la neuromatrice de la douleur par l'observation de la douleur d'autrui (résonance neurale).	Plusieurs études ont montré que la psychopathie se caractérise par un déficit de résonance empathique, notamment <i>via</i> une moindre réponse hémodynamique au niveau de l'amygdale (e.g., Blair, 2007).
Conscience de soi	Permet d'attribuer à autrui des états mentaux et affectifs, et d'accéder à la connaissance que le soi et autrui sont similaires mais distincts.	La confusion soi/autrui provoque une activation plus extensive de la neuromatrice de la douleur (cortex somatosensoriel secondaire et insula).
Flexibilité mentale et Autorégulation	Autorisent l'inhibition de la perspective personnelle, l'évaluation de la perspective d'autrui et le maintien d'une distinction soi/autrui claire et consciente.	Adopter la perspective d'autrui s'accompagne d'activations additionnelles au niveau de certaines aires corticales impliquées dans la régulation émotionnelle et les fonctions exécutives (e.g., cortex préfrontal dorsolatéral).

Contexte et similarité

L'automaticité de la résonance empathique a été remise en question par un certain nombre de travaux (e.g., Hein & Singer, 2008 ; Singer & Lamm, 2009). Des études suggèrent en effet que l'empathie pour la douleur est influencée par la manière dont les observateurs conceptualisent à la fois la situation et la personne en détresse. Ainsi, un contexte compétitif (*vs* coopératif) semble avoir un impact sur les mécanismes de résonance (Englis et al., 1982 ; Lanzetta & Englis, 1989 ; Singer et al., 2006). Par exemple, Lanzetta et Englis (1989) ont montré qu'un contexte de compétition favorise le contre-mimétisme (la douleur d'autrui peut provoquer un sourire) et la « contre-empathie » (la détresse d'autrui génère des sentiments positifs). Plusieurs études IRMf ont également montré que les réponses neurales à la douleur d'autrui sont influencées par la similarité et la familiarité avec autrui. Par exemple, une moindre activité au niveau du cortex cingulaire médian antérieur a été enregistrée quand les participants étaient exposés à une cible stigmatisée (*vs* non stigmatisée ; Decety, Echos, & Correll, 2009). Plusieurs études en imagerie cérébrale ont trouvé une moindre activité du CCA et de l'insula quand les participants étaient exposés à des personnes en souffrance appartenant à d'autres groupes ethniques (e.g., Xu, Zuo, Wang, & Han, 2009) et une plus forte activité du CPF médian face à des personnes en souffrance du même groupe ethnique (e.g., Mathur, Harada, Lipke, & Chiao, 2010).

1.2.3. Première approche de l'OD par les neurosciences sociales

Partant des travaux montrant que l'empathie est modulée par le contexte, Cheetham et al. (2009) ont posé la question de la « nature » de l'expérience subjective des participants aux études sur l'OD et ont proposé que le dilemme social induit par le paradigme de Milgram pourrait favoriser l'expérience de la détresse personnelle. Les auteurs ont eu recours à une version virtuelle du paradigme de Milgram et à l'IRMf afin d'enregistrer l'activité neurale

des participants pendant la procédure d'obéissance. Avant d'entrer dans le scanner, l'expérimentateur expliquait aux participants qu'ils allaient participer à un test devant évaluer les capacités d'apprentissage d'un avatar ayant appris une liste de paires de mots. Pour chaque essai, le mot stimulus et le mot associé étaient présentés au-dessus de l'avatar (*cf.* figure 12). Les instructions précisait qu'un choc devait être administré à l'avatar en cas d'erreur, et ce dans une fenêtre temporelle de 6s après la réponse de l'avatar (en appuyant sur un bouton avec l'index de la main droite ; en cas de bonne réponse, les participants devaient appuyer sur un autre bouton avec le même doigt). Les participants étaient également informés du fait que l'intensité du choc administré serait affichée au début de chaque essai (75, 150 ou 225 volts).

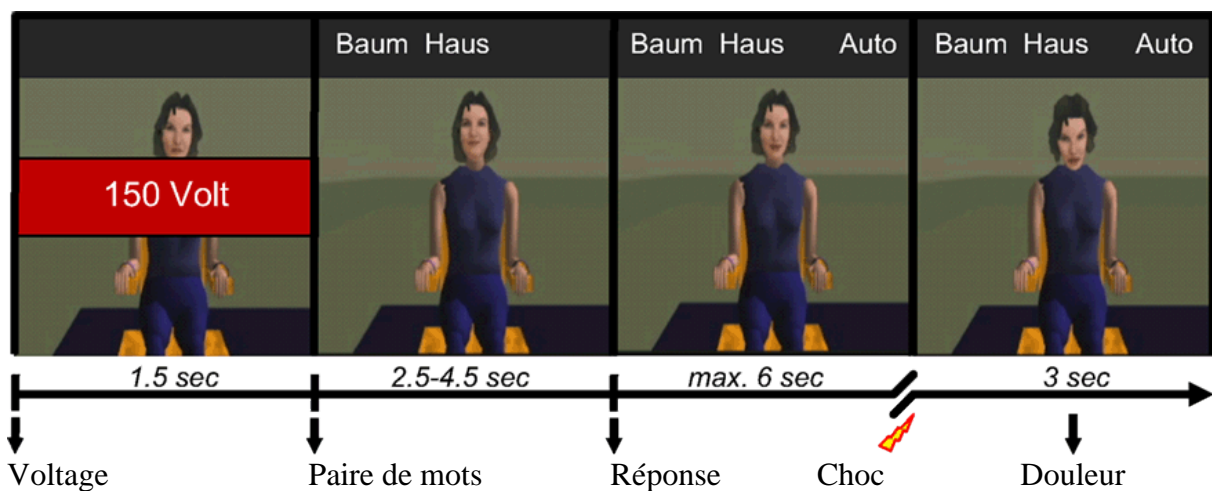


Figure 12. Déroulement temporel d'un essai avec choc électrique. Chaque essai commence par l'affichage de l'intensité du choc à administrer en cas d'erreur. Le mot stimulus (e.g., Baum ; fr. « arbre ») et le mot associé (e.g., Haus ; fr. « maison ») apparaissent ensuite au-dessus de l'avatar. Les participants sont informés du fait que l'avatar ne peut voir que le mot stimulus. Dans les essais avec chocs, l'avatar est vu considérant le mot puis donnant une mauvaise réponse (e.g., Auto ; fr. « voiture »). Les participants ont alors 6s pour appuyer sur un bouton permettant de délivrer un choc à l'avatar. Si ce délai est dépassé, un message apparaît demandant l'administration du choc (1.5s). L'événement critique de l'expérience est l'observation de la douleur de l'avatar. Dans les essais sans chocs, les participants observent simplement l'avatar donnant la bonne réponse. Les deux types d'essais se concluent par un écran blanc avec une croix de fixation (6-8s). Pour chaque essai, la durée des événements est variable et déterminée de manière à susciter un sentiment d'interaction réelle avec l'avatar. (Cheetham et al., 2009)

La procédure était divisée en trois blocs de 25 essais (7 essais avec chocs et 18 essais sans chocs ; 6 chocs pour chaque niveau d'intensité). La durée des différents événements constituant un essai était déterminée de manière à maximiser la sensation d'une interaction réelle entre le participant et l'avatar. La durée totale de la procédure d'obéissance était de 35mn. Conformément à l'hypothèse initiale des auteurs, les résultats montrent que l'observation de la douleur de la victime induit une augmentation de l'activité hémodynamique au niveau des aires cérébrales impliquées dans le traitement et l'évaluation de l'information émotionnelle comme l'AMY droite (*cf.* figure 13), mais pas au niveau des aires impliquées dans l'autorégulation comme le CCA. Ce *pattern* d'activation suggère ainsi un état de détresse personnelle chez les participants. De ce fait, Cheetham et al. (2009) ont proposé que l'obéissance dans le paradigme de Milgram puisse être (en partie) la conséquence d'un déficit de régulation de la détresse induite par l'observation de la douleur de la victime.

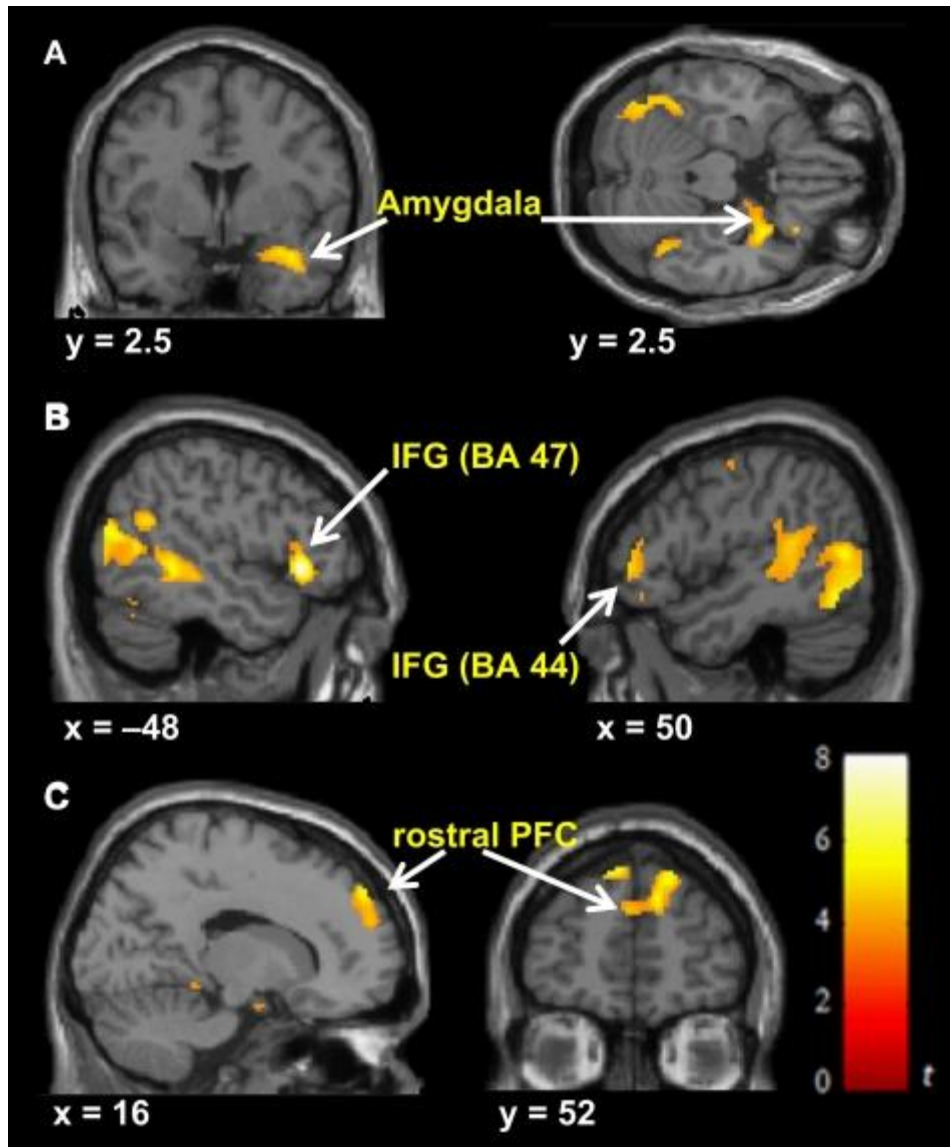


Figure 13. Réponses hémodynamiques significatives pendant l'observation de la douleur de l'avatar. Les résultats montrent des différences d'activation cérébrale entre les essais avec chocs et les essais sans chocs au niveau de l'amygdale droite et des aires périamygdaliennes (A), du gyrus frontal inférieur (B), et du cortex préfrontal rostral (C). (Cheetham et al., 2009)

Cheetham et al. (2009) ont proposé que leurs résultats offrent un « point de départ » pour une investigation plus précise du rôle des mécanismes d'autorégulation dans l'OD. Il s'agit précisément du but que nous nous sommes fixé pour le présent travail doctoral. Si la situation d'OD favorise l'expérience de la détresse personnelle, il est raisonnable de faire l'hypothèse que les variations dans la vulnérabilité au stress pourraient sous-tendre les différences interindividuelles dans l'OD. La partie suivante sera consacrée aux travaux sur l'autorégulation. Nous nous appliquerons dans un premier temps à clarifier la notion d'autorégulation à partir d'une distinction entre « réponse émotionnelle primaire » et « réponse émotionnelle secondaire ». Nous aborderons dans un second temps la question de l'influence des caractéristiques des individus et du contexte sur l'autorégulation.

1.2.4. Mécanismes d'autorégulation

Réponses émotionnelles primaires vs secondaires

Une difficulté rencontrée dans la recherche sur les émotions est la distinction entre régulation émotionnelle et d'autres formes de traitement émotionnel. Les mécanismes d'autorégulation pouvant intervenir très rapidement, des auteurs se sont focalisés sur le déroulement temporel de la réponse émotionnelle en distinguant « réponse émotionnelle primaire » et « réponse émotionnelle secondaire » (e.g., Baumann, Kaschel, & Kuhl, 2007 ; Koole, 2009 ; Lazarus, 1991). Les réponses primaires réfèrent aux réponses automatiques à des stimuli émotionnels, et reflètent la sensibilité/réactivité émotionnelle des individus. Les réponses secondaires sont guidées par les mécanismes d'autorégulation (e.g., Baumann et al., 2007). La figure 14 illustre le déroulement temporel d'une réponse émotionnelle. La réponse primaire est représentée par le gradient d'entrée (courbe ascendante) et la réponse secondaire par le gradient de sortie (courbe descendante). Si l'autorégulation désigne la facilité avec laquelle un individu peut quitter un état émotionnel spécifique, cela ne signifie pas que toutes

les stratégies d'autorégulation visent à accélérer le processus de sortie. Il est en effet possible de distinguer des stratégies inhibitrices visant à diminuer l'intensité de la réponse primaire (« down-regulation »), des stratégies excitatrices visant à augmenter cette intensité (« up-regulation ») et des stratégies visant à maintenir cette intensité sur une certaine durée (« maintenance »). Le point commun des stratégies d'autorégulation est l'altération de la courbe du gradient de sortie et ainsi la détermination de la durée de la réponse émotionnelle (Koole, Van Dillen, & Sheppes, 2013).

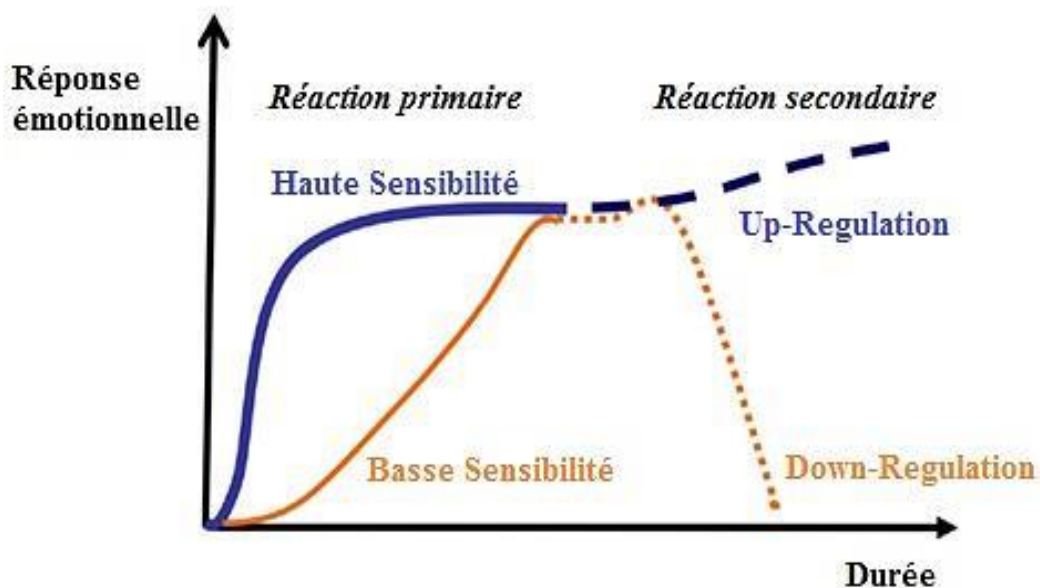


Figure 14. Modèle de la sensibilité émotionnelle vs régulation émotionnelle. « Up-Regulation » : stratégies excitatrices. « Down-Regulation » : stratégies inhibitrices. (Koole, 2009)

Les mécanismes d'autorégulation peuvent également intervenir de manière « proactive », quand par exemple un individu anticipe l'apparition d'un stimulus aversif qui devrait provoquer une réponse émotionnelle non désirée (e.g., Aspinwall & Taylor, 1997). L'intervention des mécanismes d'autorégulation précède alors la réponse émotionnelle primaire. Mais la littérature montre que l'anticipation d'une expérience émotionnelle spécifique (e.g., peur) provoque la simulation partielle de celle-ci (e.g., activation de l'AMY ;

Niedenthal, 2007). Ainsi, la réactivité émotionnelle intervient même durant l'anticipation d'une émotion non désirée.

La réactivité émotionnelle et les stratégies d'autorégulation mises en œuvre sont déterminées par la nature des stimuli, les caractéristiques de l'individu et les caractéristiques du contexte. Notamment, la recherche montre qu'il existe d'importantes différences interindividuelles dans les traits physiologiques associés à la réactivité émotionnelle et dans ceux reliés à l'autorégulation (e.g., Derryberry, Reed, & Pilkenton-Taylor, 2003). Nous aborderons dans la partie suivante la question des traits liés à l'autorégulation et la question des caractéristiques du contexte.

Notion de contrôle volontaire

Eisenberg et ses collaborateurs ont proposé qu'un état d'hyper-activation physiologique (« empathic overarousal » ; Eisenberg & Fabes, 1992) devrait favoriser une focalisation sur soi (détresse personnelle) et motiver l'adoption de comportements d'évitement. À l'inverse, un niveau optimal d'activation physiologique devrait induire une focalisation sur autrui (sympathie) et motiver l'adoption d'un comportement d'aide. Nous retrouvons ici la « loi Yerkes-Dodson » (Yerkes & Dodson, 1908 ; voir aussi Eysenck, 1955 ; Young, 1936) proposant une relation non linéaire entre activation physiologique et performance (*cf.* figure 15). Cette relation non linéaire a été vérifiée dans différents domaines du comportement humain (e.g., Anderson, 1994 ; Bregman & McAllister 1982 ; Dickman, 2002).

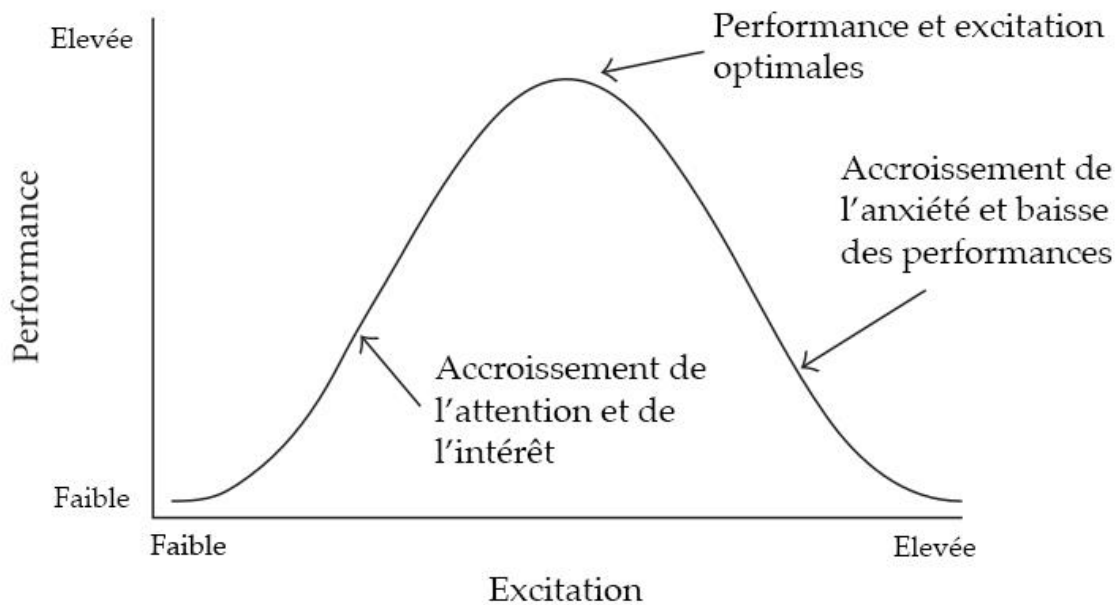


Figure 15. Illustration de la loi Yerkes-Dodson (1908) proposant une relation non linéaire entre niveau d'activation physiologique et performance.

Eisenberg et ses collaborateurs ont articulé leurs travaux autour de la notion de contrôle volontaire (« effortful control » : EC)³. L'EC a été défini comme « l'efficiace de l'attention exécutive – incluant la capacité à inhiber une réponse dominante et/ou à activer une réponse sous-dominante » (Rothbart & Bates, 2006). L'attention exécutive, qui inclut la capacité à déplacer ou focaliser volontairement son attention selon les besoins et à intégrer l'information communiquée par autrui, est un élément central de l'EC. Elle affecte à la fois le contrôle inhibiteur (capacité à supprimer les tendances à l'approche), l'activation contrôlée (capacité à réaliser une action en dépit d'une forte tendance à l'éviter), la planification et

³ Plusieurs domaines de la psychologie expérimentale se sont intéressés à l'autorégulation mais en s'appuyant sur des construits théoriques différents. L'approche développementale privilégie souvent le modèle psychobiologique du tempérament et la notion d'« effortful control » (Rothbart, Derryberry, & Posner, 1994 ; Rothbart, Ellis, & Posner, 2011 ; Rueda, Posner, & Rothbart, 2005), alors que les neurosciences cognitives abordent souvent l'autorégulation du point de vue des fonctions exécutives (Blair & Ursache, 2011 ; Gyurak et al., 2009). Du fait de l'important chevauchement conceptuel entre ces deux notions, des auteurs ont appelé à une approche unifiée pour l'étude de l'autorégulation (Zhou, Chen, & Main, 2012). De manière consistante, des études récentes suggèrent que « effortful control » et fonctions exécutives sont des construits qui se recouvrent fortement tant du point de vue fonctionnel que neurobiologique (e.g., Bridgett et al., 2013).

l'intégration de l'information. Ainsi, l'EC renvoie à un ensemble de caractéristiques ou compétences sous-tendant les différences interindividuelles dans la régulation de la réactivité émotionnelle et comportementale (e.g., Rothbart & Bates, 2006). Les études sur les substrats neuronaux de l'EC montrent un rôle important du CPF médian et du gyrus cingulaire antérieur (régions impliquées dans la régulation émotionnelle), en combinaison avec des aires du CPF latéral (région impliquée dans les fonctions exécutives ; e.g., Posner & Rothbart, 2007). Selon Eisenberg et ses collaborateurs, les individus possédant de fortes compétences d'EC devraient être moins vulnérables à la détresse personnelle et plus disposés à venir en aide à autrui en détresse. Plusieurs études ont testé cette hypothèse en utilisant diverses méthodes de mesure de l'EC (questionnaires et échelles, mesures comportementales incluant des tâches de contrôle attentionnel ou de persistance). D'abord, les auteurs ont trouvé que la détresse personnelle est associée à des réponses physiologiques plus fortes que la sympathie, et à l'adoption d'un comportement d'évitement à l'égard de la personne en détresse (pour une revue, voir Eisenberg, Valiente, & Champion, 2004 ; Eisenberg, Fabes, & Spinrad, 2006). L'EC apparaît positivement corrélé à la sympathie et négativement corrélé à la détresse personnelle chez les enfants et les adolescents (e.g., Eisenberg, Fabes, Murphy et al., 1996 ; Murphy et al., 1999 ; Valiente et al., 2004). De manière générale, ces résultats supportent l'idée voulant qu'une régulation efficace de la détresse induite par la résonance empathique favorise des réponses affectives-motivationnelles orientées vers autrui. Une limitation est le très faible nombre d'études ayant eu recours à des mesures comportementales.

Stratégies d'autorégulation basées sur l'évitement

La recherche sur la régulation émotionnelle suggère que la détresse personnelle pourrait affecter le comportement social *via* le recours à des stratégies d'autorégulation basées sur l'évitement, une réponse courante aux expériences émotionnelles aversives (e.g., Barlow,

2002 ; Gratz & Roemer, 2004). Ces stratégies incluent la distraction, l'inhibition des réponses internes, la suppression des expressions faciales émotionnelles et les comportements d'évitement (e.g. Gamez, Chmielewski, Kotov, Ruggero, & Watson, 2011 ; Hayes, Strosahl, & Wilson, 1999). L'évitement permet de réduire l'intensité d'une expérience émotionnelle à court-terme, mais est associé à un effet « rebond » d'intensité égale (e.g., Gross, 2013 ; Webb, Miles, & Sheeran, 2012).

Des auteurs ont proposé que l'AMY soit une structure cérébrale clé pour l'évaluation de la signification émotionnelle des stimuli, mais aussi dans la détection de stimuli sociaux ambigus dans l'environnement (e.g., Adams, Gordon, Baird, Ambady, & Kleck, 2003 ; LeDoux, 1996 ; Niedenthal, Mermillod, Maringer, & Hess, 2010). De manière cohérente, de nombreuses études IRMf ont montré une augmentation de l'activité de l'AMY chez les individus exposés à des stimuli menaçants (e.g., Whalen et al., 1998 ; Lane et al., 1997 ; Zald & Pardo, 1997 ; Morris et al., 1996, 1999). En revanche, des études ont montré une diminution de l'activité de l'AMY durant l'expérience et l'observation de la douleur d'autrui, associée à une diminution de l'activité sympathique (e.g., Becerra, Breiter, Wise, Gonzalez, & Borsook, 2001 ; Petrovic, Carlsson, Petersson, Hansson, & Ingvar, 2004 ; Petrovic, Ingvar, Stone-Elander, Petersson, & Hansson, 1999 ; Simon, Craig, Miltner, & Rainville 2006). Petrovic et al. (2004) ont proposé que ce phénomène de « désactivation » de l'AMY reflète une stratégie cognitive d'adaptation à des événements douloureux mais inévitables compte-tenu de la situation expérimentale.

Nous avons vu dans la partie dédiée aux neurosciences de la morale qu'une moindre activité amygdalienne est associée à une moindre aversion pour l'atteinte à autrui (e.g., Blair, 2007). L'inhibition des réponses internes comme stratégie d'autorégulation pourrait diminuer la probabilité d'une réponse prosociale face à autrui en détresse.

Notion de réévaluation cognitive

Goubert et ses collaborateurs (e.g., Goubert et al., 2005 ; Goubert, Craig, & Buysse, 2009) ont proposé que les personnes en état de détresse personnelle puissent être motivées à réévaluer la douleur d'autrui dans le sens d'une sous-estimation de son intensité, afin de maintenir leur inconfort à un niveau acceptable. La régulation émotionnelle par la réévaluation cognitive a suscité un grand nombre de travaux en psychologie et en neurosciences durant les dix dernières années. De manière générale, la réévaluation cognitive est une stratégie utilisée pour modifier l'impact émotionnel d'une situation ou d'un événement en réinterprétant sa signification (e.g., Ertl et al., 2013). Un certain nombre d'études s'est intéressé aux stratégies de réévaluation cognitive appliquées aux stimuli aversifs (e.g., Banks et al., 2007 ; Kalisch et al., 2005 ; Kanske et al., 2011 ; Lamm, Batson, & Decety, 2007 ; Ochsner et al., 2002, 2004 ; Schulze et al., 2011). Lorsqu'il est demandé aux participants de réinterpréter une scène de manière à la rendre moins négative pour eux, les résultats montrent un rôle important de l'interaction CPF-AMY dans la régulation émotionnelle. Des études en imagerie cérébrale ont mis en évidence une corrélation négative entre CPF et AMY, de même qu'une influence inhibitrice du CPF sur l'AMY et l'insula durant des tâches de réévaluation cognitive, dont la conséquence est une diminution de l'activation sympathique (*i.e.*, stress physiologique) et du sentiment d'anxiété (Delgado et al., 2008 ; Hariri et al., 2003 ; Wager et al., 2008). La recherche animale a mis en évidence le rôle important des oscillations neurales dans la gamme de fréquence thêta (3-8Hz) dans l'interaction CPF-AMY (e.g., Lesting et al., 2011 ; Narayanan et al., 2011). Dans une récente étude EEG, Ertl et al. (2013) ont utilisé des stimuli visuels aversifs afin d'induire des réactions émotionnelles négatives chez des participants qui avaient pour consigne de diminuer, augmenter ou maintenir l'émotion induite. Les auteurs ont observé (i) une augmentation de la puissance thêta au niveau du CPF dans la condition « diminution de la

réponse émotionnelle » (*vs* maintien), (ii) une corrélation positive entre puissance thêta et diminution des ressentis d'aversion. Ces résultats sont cohérents avec les travaux antérieurs ayant mis en évidence l'importance des ondes thêta dans le traitement des stimuli négatifs et dans la régulation émotionnelle (e.g., Aftanas & Golocheikine, 2001 ; Knyazev, Bocharov, Slobodskaya, & Ryabichenko, 2008).

Selon Goubert et ses collaborateurs, la réévaluation de la douleur d'autrui dans le sens d'une sous-estimation de son intensité peut s'avérer adaptative dans certains contextes. Un exemple est celui des professionnels de la santé exposés de manière chronique à des situations de détresse parfois intense, pour qui une évaluation efficace de l'intensité de la douleur d'autrui peut représenter un coût physiologique et émotionnel très important (comme l'atteste la prévalence du « burn out » dans certaines professions de la santé ; Palm, Polusny, & Follette, 2004). Pour ces professionnels, une sous-estimation relative favorisera l'initiation d'actions ajustées à la situation d'autrui (Goubert et al., 2009). En revanche, les auteurs ont également signalé que cette sous-estimation peut compromettre l'adoption d'un comportement prosocial chez les individus tout-venant.

Nous avons vu que l'observation de la douleur d'autrui provoque des réactions affectives-motivationnelles associées à des niveaux variables d'excitation physiologique. Notamment, un état d'hyperexcitation physiologique induit une focalisation sur soi (détresse personnelle) et peut motiver l'adoption de stratégies d'autorégulation incompatibles avec une réponse prosociale. L'autorégulation dépendrait de manière critique de la « flexibilité physiologique » d'un individu, c'est-à-dire de la capacité d'un individu à ajuster son niveau d'excitation physiologique à un instant précis (e.g., Gross, 1998). Le SNA est une composante clé dans la détermination du niveau d'excitation physiologique en réponse à des stimuli émotionnels. Les travaux initiés par Porges dans les années 1990 sur le rôle du SNA dans la

détermination des réactions aux stimuli externes ont permis de préciser la nature et les causes de la vulnérabilité au stress. Nous présenterons ci-dessous la théorie polyvagale mettant en avant le rôle de la régulation de la fonction cardiaque par le nerf vague dans la sélection des réponses comportementales aux stimuli externes. Nous concluons avec une brève présentation du modèle d'intégration neuroviscérale (Thayer & Lane, 2000 ; Thayer, Hansen, Saus-Rose, & Johnsen, 2009), modèle complémentaire à la théorie polyvagale portant sur le rôle des compétences attentionnelles dans la régulation du métabolisme.

1.2.5. Théorie polyvagale

1.2.5.1 Stress et homéostasie

Le SNA régule de manière continue l'homéostasie viscérale. Traditionnellement divisé en deux grandes branches, le système nerveux parasympathique (SNP) et le système nerveux sympathique (SNS), le SNA assure la répartition des ressources de l'organisme pour satisfaire à la fois les demandes internes (organisme) et externes (environnement). La balance entre les besoins internes et externes peut servir de base à la définition du stress et de l'homéostasie. Selon Porges (1992), stress et homéostasie sont interdépendants : l'homéostasie reflète la régulation dynamique du « milieu interne » par le SNP, et le stress reflète la subjugation des besoins internes par le SNS en réponse aux besoins externes. L'homéostasie correspond donc à un état autonome favorisant les besoins viscéraux en l'absence de défi externe. Elle se caractérise par une forte influence parasympathique. Par contraste, le stress est un état autonome reflétant une perturbation de l'homéostasie. Il correspond à une réponse à un défi externe et se caractérise par une diminution de l'influence parasympathique. Dans cette perspective, la mesure de la force de l'influence parasympathique en l'absence d'une demande externe permet de mesurer la vulnérabilité au stress d'un individu.

La théorie polyvagale (Porges, 1995, 1998, 2001, 2003, 2007) précise un ensemble de processus autonomes impliqués dans la détermination des réponses émotionnelles et comportementales adaptatives face aux défis posés par notre environnement. Notamment, cette théorie insiste sur l'importance de la régulation de la fonction cardiaque par le nerf vague dans la sélection des réponses face aux expériences stressantes.

1.2.5.2. Phylogénèse du SNA

Le SNA semble avoir évolué sous la pression de la vie en groupe (sous la pression à la fois du nombre des interactions et de leur complexité) afin de permettre des réponses rapides et flexibles aux changements imprévisibles de notre environnement social (e.g., Taborsky & Oliveira, 2012). Alors que le SNA est traditionnellement divisé en deux circuits neuronaux, la théorie polyvagale propose l'existence de trois circuits neuronaux hiérarchisés du point de vue phylogénétique et supportant des stratégies comportementales adaptatives spécifiques (e.g., Porges, 2007) : (i) le « système d'engagement social » régulé par le nerf vague myélinisé, dont la fonction est de favoriser des états de calme en inhibant l'influence sympathique sur le cœur, (ii) le système « fight-or-flight », dépendant du fonctionnement du SNS, (iii) le système d'immobilisation (composante la plus primitive du SNA) régulé par le nerf vague non myélinisé. Le tableau 8 résume les fonctions autonomes et comportementales associées à ces trois circuits neuronaux.

Tableau 8

Stades phylogénétiques du contrôle neural du système nerveux autonome selon Porges (2007)

Composantes du SNA	Origine des neurones moteurs	Fonctions comportementales	Fonctions autonomes
Nerf vague myélinisé (complexe vagal ventral)	Noyaux ambigus	<i>Engagement social et soin à autrui</i> Augmentation de la régulation du cœur et diminution de l'activité de l'axe HPA.	<i>Neuroprotection</i> Stabilisation des processus autonomes, incluant une augmentation de l'arythmie respiratoire sinusale, assurant la protection du cœur et l'augmentation de l'oxygénation du cerveau. Cette régulation de l'homéostasie viscérale favorise la socialité en prodiguant les ressources nécessaires aux interactions sociales.
Axe hypothalamo-hypophyso-surrénalien (SNS)	Moelle épinière	<i>Mobilisation</i> Adaptations actives, incluant les réponses défensives d'attaque et de fuite (« fight-or-flight »).	<i>Activation</i> Augmentation de la fréquence cardiaque, libération de glucocorticoïde et de catécholamine. Production d'énergie, incluant le glucose, et conversion de la noradrénaline en adrénaline.
Nerf vague non myélinisé (complexe vagal dorsal)	Noyau dorsal du nerf vague	<i>Immobilisation</i> Adaptations passives, incluant la thanatose et la perte de connaissance.	<i>Conservation</i> Prévalence de la bradycardie et de l'apnée, réduction de la production d'énergie.

Du point de vue fonctionnel, le nerf vague myélinisé fonctionne comme un « frein vagal » permettant l'inhibition ou la désinhibition rapide de l'influence vagale sur le cœur, de manière à pouvoir mobiliser ou calmer rapidement l'individu. Dans un contexte sûr, une augmentation de l'influence vagale sur le cœur permettra l'inhibition des mécanismes défensifs du SNS et l'adoption d'un comportement d'engagement social. En revanche, dans un contexte menaçant, la diminution de l'influence vagale sur le cœur permettra l'adoption de comportements défensifs (système de mobilisation ou d'immobilisation). Dans les situations de repos, le nerf vague myélinisé exerce une influence tonique constante sur le cœur (« tonus vagal »). Ainsi, le contrôle neural du cœur serait un fondement de la flexibilité physiologique.

Principe jacksonien de dissolution

Jackson (1958) a proposé que les structures cérébrales les plus récentes (phylogénétiquement) inhibent les structures les plus anciennes, et lorsque les structures les plus récentes sont compromises les plus anciennes prendraient le relais. De manière cohérente avec ce « principe jacksonien de dissolution », la théorie polyvagale précise que face à un défi environnemental chaque circuit neural sera recruté de manière séquentielle, l'échec du circuit le plus récent entraînant la mobilisation du circuit précédant du point de vue phylogénétique (engagement social, puis mobilisation, et enfin immobilisation). Porges (1996) a insisté sur le fait que l'homéostasie viscérale et l'engagement social promu par le circuit le plus récent sont incompatibles avec les états neurophysiologiques et les stratégies comportementales régulés par les circuits supportant les stratégies défensives. Ainsi, l'activation du système d'engagement social nécessite l'inhibition des systèmes défensifs (et inversement).

Neuroception du danger

L'évaluation du risque est déterminante pour l'adoption d'un comportement prosocial. Des travaux suggèrent que l'évolution ait doté l'humain d'un biais vers la détection de la menace (*i.e.*, asymétrie attentionnelle en faveur de la détection du danger ; e.g., Neuberg, Kenrick, & Schaller, 2011). La valeur adaptative de ce biais réside dans la capacité accrue d'un organisme à répondre rapidement aux stimuli potentiellement nocifs (e.g., Johnson, Blumstein, Fowler, & Haselton, 2013). Ce biais attentionnel serait sous-tendu par un stress maintenu constant rendant l'organisme plus vulnérable aux stressseurs (e.g., Sandi & Haller, 2015). La « neuroception » désigne un processus neural évaluant le risque de manière continue et modulant l'activité vagale *via* certaines structures cérébrales supérieures (e.g., Porges, 2007). Dans un environnement perçu comme sûr, le mode de fonctionnement par défaut de détection de la menace peut être inhibé, réduisant ainsi le niveau de stress et

permettant l'adoption de comportements prosociaux. Une neuroception défailante induisant une surestimation chronique de la menace aura pour conséquence un défaut d'inhibition des mécanismes défensifs du SNS dans un contexte sûr.

1.2.5.3. Mesure du tonus vagal

Le tonus vagal est considéré comme un indicateur robuste de la flexibilité autonome et des compétences à l'autorégulation, ou encore comme un biomarqueur de la vulnérabilité au stress (e.g., Porges, 2007). La mesure du tonus vagal repose sur les principes suivants : (i) le nerf vague contrôle de manière dynamique la fréquence cardiaque qui varie continuellement en réponse aux stimuli internes et externes, de manière à adapter l'organisme à son environnement tout en assurant l'homéostasie viscérale, (ii) l'analyse de cette variabilité de la fréquence cardiaque (« heart rate variability » : HRV) repose sur la mesure des intervalles R-R consécutifs du rythme cardiaque à partir d'un électrocardiogramme (*cf.* figure 17). L'analyse de la HRV par les méthodes temporelle et spectrale permet d'obtenir des indices reflétant la force de l'influence vagale (parasymphatique).

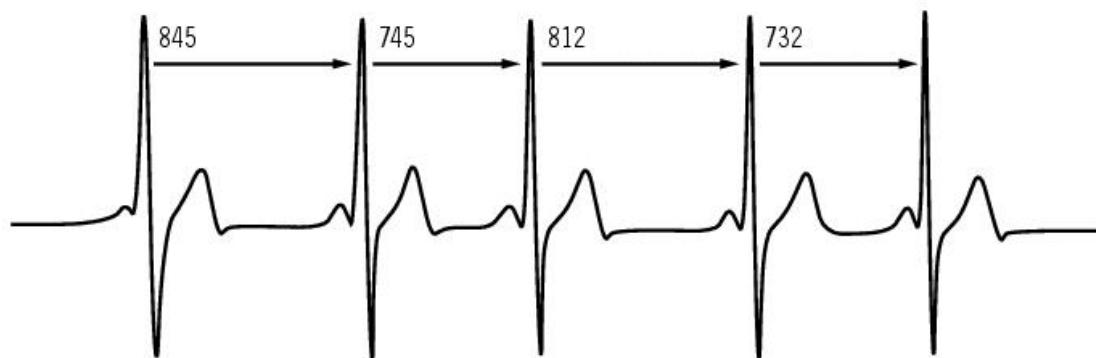


Figure 16. Le complexe QRS réfère à 3 ondes accolées correspondant à la dépolarisation des ventricules. L'onde Q est la première onde négative, l'onde R la première onde positive du complexe, et l'onde S la première onde négative après l'onde R. L'analyse de la HRV repose sur la mesure des intervalles R-R consécutifs.

Domaine temporel

L'analyse de la HRV peut être effectuée dans le domaine temporel en mesurant les durées entre les intervalles R-R et leur déviation standard. Cette analyse est influencée à la fois par les activités parasympathique et sympathique. Plusieurs indices peuvent être calculés à partir des différences entre les intervalles R-R, dont le RMSSD (« root mean squared successive differences » - racine carrée de la moyenne des différences au carré entre les intervalles R-R successifs). Cet indice mesure essentiellement l'évolution moyenne des intervalles R-R. Il reflète également la variabilité de haute fréquence principalement d'origine parasympathique, modulée par la respiration. Le RMSSD est recommandé pour évaluer le tonus parasympathique (« Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology », 1996). Sa diminution est associée à une moindre homéostasie viscérale et une plus grande vulnérabilité au stress (Porges, 2007).

Domaine fréquentiel

L'analyse fréquentielle est le calcul du contenu spectral d'une série chronologique des intervalles R-R consécutifs. L'analyse spectrale permet d'identifier des oscillations périodiques en séparant les composantes en hautes fréquences (entre 0.15 et 0.45 Hz) et basses fréquences (entre 0.04 et 0.15 Hz). Les hautes fréquences sont un indicateur de l'activité parasympathique, alors que les basses fréquences traduiraient principalement l'activité sympathique (e.g., Pagani et al., 1986). Des études ont montré que les paramètres RMSSD et variabilité cardiaque de haute fréquence (HF-HRV) sont positivement corrélés (e.g., Otzenberger et al., 1998).

1.2.5.4. Évidence empirique

Suivant la théorie polyvagale, l'amplitude de la variabilité de la fréquence cardiaque à l'état de repos (« HRV basale » : HRVb) devrait permettre de prédire les compétences à l'autorégulation et les tendances prosociales des individus. De manière consistante, une moindre HRVb a été associée à des réponses émotionnelles plus négatives en situation de stress (Fabes & Eisenberg, 1997), à la détresse personnelle en réponse à la douleur d'autrui (Fabes et al., 1993), à l'anxiété sociale et au stress psychosocial (e.g., Thayer & Lane, 2007), et à l'adoption de comportements défensifs comme l'évitement et l'agression dans les situations menaçantes, incertaines ou nouvelles (e.g., Movius & Allen, 2005). Une récente étude portant sur la relation entre tonus vagal et stratégie spontanée d'autorégulation face à des stimuli aversifs a mis en évidence qu'une moindre HRVb est associée à une plus forte tendance à l'évitement (e.g., inhibition des réponses internes, distraction ; Aldao et al., 2015). Par contraste, des études ont montré qu'une plus forte HRVb est associée à une meilleure régulation des réponses psychophysiologiques en situation de stress et aux sentiments prosociaux comme la sympathie et la compassion (e.g., Fabes & Eisenberg, 1997 ; Stellar, Cohen, Oveis, & Keltner, 2015). Muhtadie, Koslov, Akinola, et Mendes (2015) ont proposé que la réactivité vagale (*i.e.*, variation de la HF-HRV selon la demande de l'environnement) soit un biomarqueur spécifique de la sensibilité sociale. Des travaux ont montré qu'une demande cognitive, sociale, émotionnelle ou sensorielle provoque une diminution de la HF-HRV relativement à l'état de repos (e.g., Van Roon, Mulder, Althaus, & Mulder, 2004) et que la magnitude de cette réactivité vagale est un prédicteur de la prosocialité (e.g., Akinola & Mendes, 2014 ; Obradović, Bush, Stamperdahl, Adler, & Boyce, 2010). En situation d'interaction sociale, Muhtadie et al. (2015) ont mis en évidence que plus la réactivité vagale des individus est importante, plus ces derniers sont sensibles et réactifs au comportement d'autrui.

1.2.6. Modèle d'intégration neuroviscérale

De manière générale, le modèle d'intégration neuroviscérale propose que les réseaux neuronaux impliqués dans la régulation attentionnelle (e.g., CPFVM, COF, AMY) soient impliqués dans le contrôle de l'activité cardiaque *via* le nerf vague (Thayer & Lane, 2000, 2009 ; Park & Thayer, 2014). Ce modèle propose également que le stress soit un mode de fonctionnement par défaut assuré par les circuits sympatho-exciteurs (notamment l'AMY) orientant l'organisme vers la détection de la menace. En l'absence de menace, l'influence inhibitrice du CPF sur les circuits sympatho-exciteurs faciliterait les tendances prosociales. Cette influence inhibitrice modulerait la variabilité de la fréquence cardiaque *via* la branche myélinisée du nerf vague. Dans un contexte menaçant, la diminution de cette influence inhibitrice favoriserait la vigilance par défaut à la menace et les comportements défensifs comme l'évitement ou l'agression (e.g., Park, Vasey, Van Bavel, & Thayer, 2013). La figure 16 illustre le contrôle neural de la fonction cardiaque.

De manière cohérente, des études en imagerie cérébrale ont montré qu'une diminution de l'activité du CPF est responsable d'une hyperactivité sous-corticale, une hypervigilance et l'expression rigide de comportements défensifs (e.g., Thayer et al., 2009 ; Park et al., 2013). La recherche a également montré que le fonctionnement des circuits préfrontaux et sous-corticaux influence l'activité cardiaque *via* le nerf vague. Par exemple, des études ont montré qu'une plus forte HRVb est associée à des performances plus élevées à des tâches cognitives impliquant les fonctions exécutives, une plus grande flexibilité attentionnelle et émotionnelle, alors qu'une moindre HRVb est associée à une hypoactivité préfrontale, une hyperactivité des structures sous-corticales, et une moindre flexibilité attentionnelle et émotionnelle (voir Thayer, Ahs, Fredrikson, Sollers, & Wager, 2012). Des études ont également démontré qu'une surestimation chronique de la menace est associée à une hyperactivité amygdalienne,

un plus fort biais de négativité (*i.e.*, tendance à allouer plus de ressources au traitement des stimuli négatifs) et une moindre tendance à l'approche vis-à-vis des stimuli positifs nouveaux (e.g., Shook, Pena, Fazio, Sollers, & Thayer, 2007 ; Thayer et al., 2012). Ainsi, la HRVb reflèterait l'inhibition du stress par le CPF. La régulation attentionnelle serait déterminante pour l'adoption d'un comportement prosocial : la capacité à inhiber le mode de fonctionnement par défaut orienté vers la menace et à déplacer l'attention vers un ensemble plus large de stimuli augmenterait la flexibilité en diminuant le niveau de stress, et faciliterait les tendances prosociales (e.g., sympathie, compassion, comportement d'aide).

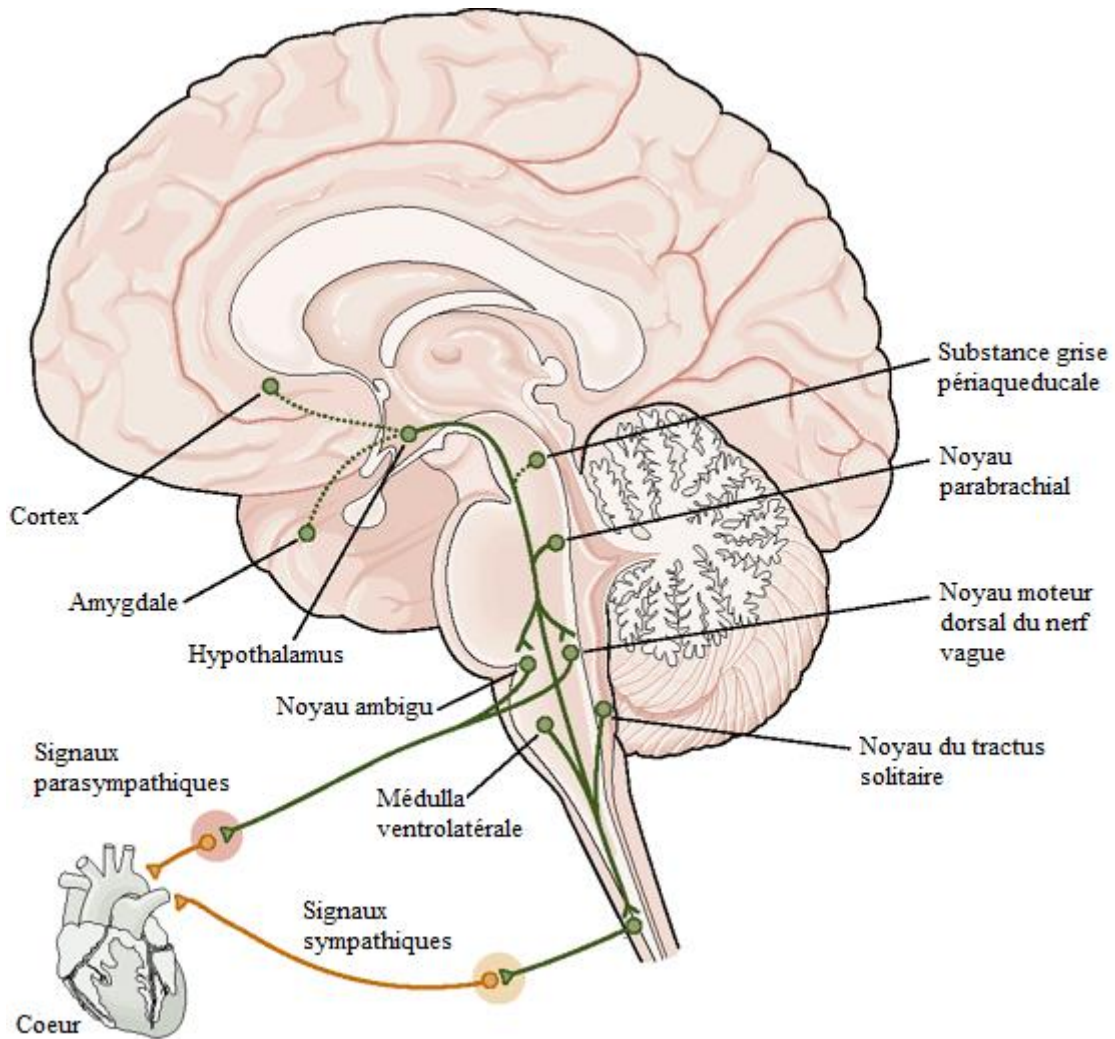


Figure 17. Contrôle neural du cœur. L'amygdale exerce une influence inhibitrice sur le noyau du tractus solitaire et une influence excitatrice sur la médulla ventrolatérale. Cette structure sous-corticale participe ainsi à la modulation des réponses du système nerveux autonome aux stimuli aversifs. Les structures préfrontales (notamment le cortex préfrontal ventromédian et le cortex orbitofrontal) exercent une influence inhibitrice sur l'amygdale. Elles participent par ce biais au contrôle de l'activité cardiaque en inhibant les réponses sympatho-excitatrices initiées par l'amygdale.

La théorie polyvagale et le modèle d'intégration neuroviscérale proposent que la flexibilité physiologique (i) soit déterminée par les interactions cœur-cerveau, (ii) sous-tende le comportement prosocial. La théorie polyvagale insiste sur le rôle du nerf vague dans la stabilisation des processus autonomes et la libération des ressources nécessaires aux interactions sociales. Le modèle d'intégration neuroviscérale intègre le rôle de l'influence inhibitrice du CPF sur les circuits sympatho-excitateurs dans le contrôle de l'activité cardiaque. L'inhibition de la vigilance par défaut assurée par l'AMY est essentielle pour permettre un déplacement de l'attention vers des stimuli autres que menaçants. Si la recherche empirique montre que le tonus vagal permet de prédire les tendances prosociales des individus, une limitation est le très faible nombre d'études ayant exploré la relation entre tonus vagal et mesures du comportement prosocial. Nous pouvons citer l'étude de Miller, Kahle et Hastings (2015) mettant en évidence une relation entre comportement altruiste et fonctionnement vagal chez les enfants. Beffara, Bret, Vermeulen et Mermillod (2016) ont quant à eux trouvé qu'une moindre HF-HRV à l'état de repos est associée à une moindre tendance à la coopération. L'état actuel des connaissances permet la formulation des deux hypothèses suivantes : (i) une moindre flexibilité physiologique devrait favoriser l'OD, (ii) l'influence de la flexibilité physiologique sur l'OD pourrait être médiatisée par l'exercice d'un contrôle inhibiteur sur les réponses émotionnelles automatiques associées à la résonance empathique. Nous terminerons cette partie consacrée aux apports des neurosciences par la présentation de travaux réalisés dans le champ émergent des neurosciences politiques. De manière intéressante, des études suggèrent qu'autoritarisme et conservatisme seraient étroitement associés à la flexibilité cognitive et au traitement de la menace.

1.2.7. Neurosciences politiques

Comme nous l'avons vu dans notre partie sur l'autoritarisme, une longue tradition dans la recherche en psychologie politique propose que l'autorégulation puisse jouer un rôle important dans la détermination des attitudes et comportements politiques : l'autoritarisme a été associé aux sentiments d'insécurité et de peur (Allport, 1954), le dogmatisme à une organisation cognitive relativement close de croyances et non-croyances (Rokeach, 1954, 1960), et le conservatisme à une susceptibilité généralisée à faire l'expérience de la peur et de l'anxiété (Wilson, 1973). Plus récemment, Jost et ses collaborateurs ont proposé que les individus puissent endosser des croyances conservatrices dans le but de diminuer leurs sentiments d'anxiété et de menace face à l'incertitude (e.g., Jost & Amodio, 2012). Plus précisément, les différences interindividuelles dans l'attrait pour le conservatisme seraient la conséquence de différences dans l'orientation psychologique envers l'incertitude (motivation épistémique), la menace (motivation existentielle) et la conformité (motivation relationnelle) : les croyances conservatrices seraient plus « attractives » pour les individus présentant (de manière chronique ou temporaire) de forts besoins de gestion de l'incertitude et de la menace (e.g., le maintien du *statu quo* préserverait les individus des incertitudes associées au changement). De manière consistante, une méta-analyse montre que les facteurs situationnels et dispositionnels associés à la gestion de l'incertitude et de la menace sont positivement corrélés au conservatisme politique (Jost, Glaser, Kruglanski, & Sulloway, 2003). Une limite à ces différentes propositions théoriques est qu'elles reposent essentiellement sur l'auto-report subjectif. De récentes études intégrant les composantes biologiques du comportement ont mis en évidence le rôle de la flexibilité cognitive et du traitement de la menace dans la détermination des attitudes et comportements politiques.

Flexibilité cognitive

Amodio et al. (2007) ont exploré la relation entre orientation politique et détection de conflit⁴. Les auteurs ont enregistré l'activité électrocorticale des participants pendant une tâche « Go/No-Go »⁵. Les résultats montrent que le conservatisme est associé à une plus grande persistance comportementale (plus d'erreurs sur les essais « No-Go ») et à une moindre réponse neurocognitive au niveau du CCA. Suivant une procédure similaire, Weissflog et al. (2010) ont répliqué ces résultats avec le RWA. Ces données sont compatibles avec l'idée voulant que le RWA et le conservatisme soient associés à une moindre flexibilité cognitive. Elles sont également cohérentes avec les travaux montrant que l'influence inhibitrice exercée par le CPF sur les circuits sympatho-excitateurs joue un rôle dans les attitudes intergroupes. Par exemple, des études ont montré que l'AMY est impliquée dans le traitement des informations associées à l'identité ethnique d'autrui, et qu'une augmentation de l'activité de l'AMY est associée à une augmentation de l'expression des préjugés racistes (e.g., Eberhardt, 2005 ; Hart et al., 2000 ; Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005 ; Phelps et al., 2000). D'autres études ont trouvé qu'une augmentation de l'activité du CPF latéral est inversement corrélée à l'activité de l'AMY en réponse à la présentation de visages de membres du hors-groupe (e.g., Cunningham et al., 2004).

⁴ La détection de conflit est un aspect important de l'autorégulation. Alors que la répétition d'une action au sein d'un environnement stable favorise un fonctionnement automatique, une situation complexe (e.g., nouvelle et dynamique) exige une reprise active du contrôle afin de sélectionner les actions les plus appropriées (e.g., Shallice, 1988). La théorie de la détection du conflit (Botvinick et al., 2001) propose que la détection de situations conflictuelles augmente le contrôle comportemental. La structure cérébrale impliquée dans la détection du conflit est le CCA.

⁵ La tâche « Go/No-Go » est souvent utilisée dans la recherche sur la détection de conflit. Le principe est le suivant : deux types de stimuli apparaissent de manière séquentielle sur un écran, un stimulus « Go » et un stimulus « No-Go ». Les participants doivent appuyer le plus rapidement possible sur une touche lorsque le stimulus « Go » apparaît. Le stimulus « Go » représentant la majorité des essais, l'inhibition du geste d'appuyer sur la touche à la présentation du stimulus « No-Go » exige une détection rapide du conflit.

Traitement de la menace

Feldman, Lavine, Lodge et Verhulst (2010) ont trouvé une forte association entre les scores de RWA des individus et la reconnaissance automatique de mots négatifs suggérant une menace à l'ordre social et à la cohésion de groupe. D'autres études ont montré que le conservatisme est associé à une plus forte tendance à percevoir des expressions faciales émotionnelles « ambiguës » comme menaçantes (Vigil, 2010), et des images négatives comme plus négatives (Hibbing et al., 2014). De manière générale, un nombre croissant d'études démontre que le conservatisme est associé à une focalisation attentionnelle sur les stimuli associés à l'incertitude et la menace (e.g., Hibbing et al., 2014). L'association entre idéologie et traitement de la menace a été examinée à un niveau physiologique par Oxley et al. (2008). Ces auteurs ont trouvé que le conservatisme est associé à des réponses électrodermales de plus grande amplitude (indice d'activation sympathique) à des images menaçantes, et à un plus fort réflexe de clignement en réponse à des stimuli sonores inattendus. Un peu plus récemment, Kanai et al. (2011) ont exploré la relation entre conservatisme et aspect volumétrique du CCA et de l'AMY. La littérature montre que les personnes possédant un CCA de moindre volume présentent une moindre capacité à maintenir un niveau optimum d'activation émotionnelle et un moindre niveau de contrôle cognitif dans les situations émotionnelles ; les personnes ayant une AMY d'un volume plus important présentent des réponses physiologiques et psychologiques plus importantes face aux événements émotionnels (Kanai et al., 2011). Conformément à leurs hypothèses, les auteurs ont trouvé que le conservatisme est associé (i) à une AMY droite d'un volume plus important, (ii) à un CCA d'un moindre volume.

Nous terminerons cette partie introductive par la présentation des enjeux éthiques et méthodologiques associés à l'étude expérimentale de l'OD. Nous aborderons dans un premier

temps l'usage de la duperie en psychologie sociale expérimentale et le problème de la suspicion des participants. Nous présenterons ensuite les alternatives actuelles à l'usage de la duperie, à savoir le jeu de rôle actif et la technologie d'environnement immersif.

CHAPITRE 3. ENJEUX ETHIQUES ET METHODOLOGIQUES DANS LA RECHERCHE SUR L'OD

Les travaux pionniers de Milgram (1963) ont suscité une très forte controverse éthique et l'évolution des principes éthiques guidant la recherche a rapidement proscrit toute possibilité de réplique. Milgram a notamment été critiqué pour son usage de la duperie et pour avoir exposé ses participants à un stress important (e.g., Baumrind, 1964 ; Fischer, 1968 ; Kaufmann, 1967 ; Mixon, 1972). Le recours à la duperie dans la recherche en psychologie sociale expérimentale a suscité des questionnements éthiques mais aussi méthodologiques : alors que la duperie est considérée comme un outil indispensable à l'étude des aspects socialement indésirables du comportement humain, un certain nombre d'études suggère que celle-ci pourrait au contraire favoriser la suspicion et les biais de demande chez les participants, réduisant alors le contrôle expérimental qu'elle est censée maximiser (e.g., Hertwig & Ortmann, 2008 ; Orne, 2002).

1.3.1. Coût méthodologique du recours à la duperie

L'usage de la duperie dans la recherche implique de communiquer de manière intentionnelle de fausses informations aux participants (e.g., mentir volontairement sur les objectifs d'une étude et sur l'identité du chercheur et du compère ; Hertwig & Ortmann, 2008 ; Hey, 1998). Deux arguments sont traditionnellement utilisés pour soutenir le recours à la duperie : (i) la duperie serait un outil indispensable au contrôle expérimental, notamment pour l'étude des aspects « socialement indésirables » du comportement comme l'obéissance, le conformisme, ou encore la discrimination (e.g., Cooper, 1976 ; Kimmel, 1996 ; Weber & Cook, 1972 ; Weiss, 2001), (ii) les effets de la suspicion sur le comportement des participants en contexte expérimental seraient négligeables (e.g., Kimmel, 1998).

Le fait que l'expérimentateur « contrôle » l'information délivrée aux participants et qu'il n'informe pas pleinement ces derniers sur ses objectifs et ses critères d'évaluation

provoquerait cependant un ensemble d'activités inférentielles et interprétatives chez les participants dans un effort de meilleure compréhension des intentions et attentes de l'expérimentateur (e.g., Riecken, 1962). Une question importante est alors de savoir si les participants suspicieux (*i.e.*, qui suspectent l'expérimentateur de mentir) sont d'autant plus motivés à réduire cette « asymétrie informationnelle ». Si tel est le cas, leur comportement devrait différer des participants non suspicieux (e.g., Hertwig & Ortmann, 2008). Nous aborderons dans un premier temps la question de la suspicion dans les expériences de Milgram, puis celle des effets comportementaux de la suspicion dans la recherche en psychologie sociale expérimentale.

Suspicion dans les expériences de Milgram

Le questionnaire adressé aux participants un an après leur participation portait en partie sur leur degré de conviction quant à la réalité des chocs. Les résultats montrent qu'environ 20% des participants se sont déclarés suspicieux (questions 3-5 : *cf.* tableau 8). Milgram (1974) précise que pour chaque expérience deux à quatre participants ont rapporté une suspicion. Estimant l'évaluation de cette suspicion trop imprécise (e.g., confusion entre suspicion réelle et rationalisation), Milgram n'a pas exclu ces participants des analyses.

Tableau 9
Taux de suspicion dans les expériences de Milgram (Milgram, 1974)

Questions	Participants désobéissants	Participants obéissants	Tous participants
1. J'étais absolument certain que l'élève recevait des chocs douloureux.	62.5% (230)	47.9% (139)	56.1% (369)
2. Bien que j'aie eu quelques doutes, je croyais que l'élève recevait des chocs.	22.6% (83)	25.9% (75)	24% (158)
3. Je n'étais pas certain que l'élève recevait des chocs.	6% (22)	6.2% (18)	6.1% (40)
4. Bien que j'aie eu quelques doutes, je ne croyais pas que l'élève recevait des chocs.	7.6% (28)	16.2% (47)	11.4% (75)
5. J'étais certain que l'élève ne recevait pas de chocs.	1.4% (5)	3.8% (11)	2.4% (16)

Recherche sur les effets de la suspicion

Afin d'évaluer les effets comportementaux de la suspicion, Hertwig et Ortmann (2008) ont examiné trois groupes d'études : (i) le premier groupe compare le comportement conformiste de participants suspicieux vs non suspicieux (catégorisation post-expérimentale), (ii) le deuxième groupe comprend des études reposant sur l'induction d'une suspicion, (iii) le troisième groupe réunit des études explorant l'influence du « passé expérimental » des participants.

(i) Effet de la suspicion auto-reportée sur le comportement conformiste

Hertwig et Ortmann (2008) se sont intéressés à quatorze études portant sur le conformisme et faisant usage de la duperie. Avec en moyenne 45.8% de participants suspicieux, les résultats des auteurs montrent que dans dix de ces études, les participants

suspicieux (vs non suspicieux) se sont révélés moins conformistes (taille d'effet de moyenne à large). Les auteurs concluent que dans la recherche sur le conformisme, la suspicion augmente significativement le risque d'erreur de type II.

(ii) Effets de l'induction expérimentale de la suspicion

Hertwig et Ortmann (2008) ont retenu huit études ayant induit de diverses manières un état de suspicion chez les participants. Les résultats suggèrent que trois types d'induction ont impacté le comportement des participants : (i) donner une information détaillée sur les objectifs de l'étude (e.g., Levy, 1967 ; Turner & Simons, 1974), (ii) annoncer le recours à la duperie et les modalités de son usage (Finney, 1987), (iii) décrire la manipulation expérimentale (Golding & Lichtenstein, 1970). En revanche, la simple annonce d'un recours à la duperie ne semble pas affecter le comportement (*i.e.*, aucune différence significative entre condition expérimentale et condition contrôle ; Allen, 1983 ; Finney, 1987 ; Wiener & Erker, 1986).

(iii) Effets du « passé expérimental » des participants

Est-ce que le fait d'avoir déjà participé à une ou plusieurs études impliquant un recours à la duperie affecte le comportement des participants ? Hertwig et Ortmann (2008) ont retenu neuf études ayant pris en compte le « passé expérimental » des participants, et les résultats suggèrent que : (i) l'expérience personnelle de la duperie affecte le comportement des participants, ce qui n'est pas le cas de la simple divulgation de la possibilité d'une duperie (Cook & Perrin, 1971), (ii) l'expérience personnelle de la duperie exacerbe l'appréhension d'être évalué (Silverman, Shulman, & Wiesenthal, 1970), (iii) l'effet de l'expérience personnelle de la duperie dépend de la nature des variables dépendantes (e.g., les données relatives aux attitudes des participants ne semblent pas être affectées ; Cook et al., 1970), (iv)

l'effet de l'expérience personnelle de la duperie semble dépendre de la similarité entre situation expérimentale présente et passée (Brock & Becker, 1966 ; Cook et al., 1970).

Afin de permettre l'étude expérimentale de l'OD sur des bases éthiques et méthodologiques optimales, des chercheurs ont travaillé à l'élaboration de paradigmes d'obéissance alternatifs reposant notamment sur le jeu de rôle actif. Aujourd'hui, un nombre important d'études suggère que son usage représente une alternative fiable et valide pour l'étude du comportement social de manière générale, et l'étude de l'OD de manière plus spécifique.

1.3.2. Jeu de rôle actif et OD

Trois types de jeu de rôle ont été utilisés dans la recherche sur l'OD (Meeus & Raaijmaker, 1995) : (i) « description-imagination » (une description détaillée du paradigme de Milgram est fournie aux participants qui doivent imaginer quel serait leur comportement), (ii) jeu de rôle non-actif (une vidéo de l'expérience est présentée et les participants doivent dire comment ils se seraient eux-mêmes comportés), (iii) jeu de rôle actif (les participants sont informés du fait que l'expérience est une simulation mais ont pour consigne de se comporter comme si la situation était réelle). La méthode « description-imagination » produit une sous-estimation chronique et majeure du taux d'obéissance (en moyenne, moins de 10% des participants déclarent aller jusqu'au bout de la procédure d'obéissance ; e.g., Freedman, 1969 ; Milgram, 1963, 1974 ; Miller, Gillen, Schenker, & Radlove, 1974 ; Mixon, 1972). La méthode « jeu de rôle non-actif » produit des taux d'obéissance plus proches de ceux obtenus par Milgram, mais toujours inférieurs (en moyenne 43% pour la condition standard de Milgram ; e.g., Bierbrauer, 1979 ; Hess, 1971 ; Mixon, 1972). En revanche, douze études reposant sur l'usage du jeu de rôle actif ont permis de reproduire les résultats de Milgram (Albrecht, 1973 ; Bierbrauer, 1979 ; Geller, 1978 ; Holland, 1969 ; Mixon, 1972 ; O'Leary,

Willis, & Tomich, 1970). Par exemple, O'Leary, Willis et Tomich (1970) ont réalisé une réplification exacte de l'expérience 2 de Milgram (condition standard) sur la base du jeu de rôle actif : les participants étaient informés avant le début de l'expérience que l'élève ne recevrait aucun choc mais se comporterait comme s'il en recevait réellement. Le tableau 9 présente les résultats des auteurs mis en regard de ceux de Milgram.

Tableau 10

Taux d'obéissance dans la condition standard de Milgram (1963) et dans l'étude d'O'Leary et al. (1970) reposant sur le jeu de rôle actif

	Milgram (1963)	O'Leary et al. (1970)
Refus de commencer	0	5
15 volts – 285 volts	0	0
300 volts	12.5	0
315 volts – 390 volts	22.5	25
450 volts	65.5	70

1.3.3. Technologie d'environnement immersif

Dans les années 1990, des chercheurs se sont intéressés à la réalité virtuelle comme outil permettant l'étude des déterminants du comportement. La finalité de la réalité virtuelle est de permettre à une personne (ou à plusieurs) une activité sensori-motrice et cognitive dans un monde artificiel, créé numériquement, pouvant simuler certains aspects du monde réel. Un environnement virtuel immersif (EVI) permet entre autres de simuler des interactions dynamiques et en temps réel entre un participant et un avatar au sein d'un environnement

artificiel reproduisant certaines situations sociales (e.g., Blascovich et al., 2002 ; Witmer & Singer, 1998).

La technologie d'environnement virtuel immersif (« Immersive Virtual Environment Technology » : IVET) présente des avantages méthodologiques importants pour la psychologie sociale. Tout d'abord, le recours à la simulation a le potentiel de permettre « l'isolation sans altération » (Dambrun & Vatiné, 2010 ; Greenwood, 1989). Ensuite, l'IVET peut grandement faciliter la réplique exacte. En effet, si la réplique est souvent considérée comme un aspect important de la démarche scientifique, celle-ci est problématique tant pour les études reposant sur la duperie (les chercheurs ne partagent pas le même environnement expérimental ; Blascovich et al., 2002) que pour celles impliquant le jeu de rôle actif (les compères impliqués ne possèdent pas les mêmes caractéristiques physiques ; Dambrun & Vatiné, 2010). En revanche, l'usage d'un même programme d'EVI par différentes équipes de recherche autorise des répliques quasi parfaites (Blascovich et al., 2002). Enfin, l'IVET peut contribuer à résoudre le problème du compromis contrôle expérimental/réalisme ordinaire qui caractérise la recherche en psychologie sociale expérimentale (Blascovich et al., 2002 ; Loomis, Blascovich, & Beall, 1999). L'usage de scénarios minimalistes (e.g., utilisation de simples vignettes écrites) augmente le contrôle expérimental mais réduit fortement le réalisme ordinaire. Or, ce dernier augmente l'engagement des participants dans la situation expérimentale et leur sensibilité aux inductions expérimentales (e.g., Aronson & Carlsmith, 1969). À l'inverse, l'usage de scénarios complexes (e.g., implication de compères entraînés et explications élaborées) augmente le réalisme ordinaire mais diminue le contrôle expérimental. Ainsi, plus le réalisme ordinaire augmente, plus le contrôle expérimental diminue (Blascovich et al., 2002). Cependant, l'usage de l'IVET permet l'immersion des participants dans des environnements

réalistes tout en permettant un contrôle quasi optimal de l'environnement expérimental et des actions s'y déroulant.

Ainsi, l'IVET permet de maximiser à la fois le réalisme ordinaire et le contrôle expérimental, et facilite la réplication exacte. Ces avantages font de l'IVET un outil à fort potentiel pour la recherche sur le comportement humain. Les études comportementales et IRMf réalisées depuis le début des années 2000 semblent confirmer que l'IVET est un outil valide et fiable pour l'étude du comportement social (e.g., Bailenson & Yee, 2005 ; Blascovich et al., 2002 ; Freeman et al., 2003, 2005 ; Schilbach et al., 2006). L'IVET a été utilisé pour l'étude des aspects à la fois normaux et pathologiques du comportement, comme l'anxiété sociale (Pertaub, Slater, & Barker, 2002), les tendances paranoïdes (Freeman et al., 2003, 2005), la facilitation/inhibition sociale (Blascovich et al., 1999 ; Bond & Titus, 1983) ou encore la comparaison sociale et le conformisme (Swinth & Blascovich, 2001). L'utilisation de l'IVET a permis entre autres la réplication de résultats classiques obtenus dans des environnements réels dans le domaine de la psychologie sociale (e.g., Blascovich et al., 1999 ; Bond & Titus, 1983). Ces travaux sont autant d'exemples de la « présence », c'est-à-dire la tendance des participants à répondre à des situations, événements et personnages virtuels comme s'ils étaient réels, et ce tant au niveau neural que comportemental (e.g., Held & Durlach, 1992 ; Sanchez-Vives & Slater, 2005 ; Schilbach et al., 2006 ; Sheridan, 1992).

« *Virtual Milgram paradigm* »

Slater et al. (2006) ont réalisé une version virtuelle du paradigme de Milgram. Le but des auteurs n'était toutefois pas d'étudier l'OD mais l'interaction humain/avatar dans une situation sociale extrême. Plus précisément, la question était de savoir dans quelle mesure les participants pourraient répondre à une situation virtuelle d'OD comme si la victime et les chocs étaient réels. Pour cette étude, les auteurs ont comparé les réponses comportementales,

physiologiques et subjectives des participants dans une situation où la victime virtuelle était visible à une situation où elle ne l'était pas. Comme dans le paradigme de Milgram, les participants devaient administrer une série de chocs électriques (x20) à l'élève (virtuel) à chaque fois que celui-ci commettait une erreur à un test de mémoire (*cf.* figure 18).

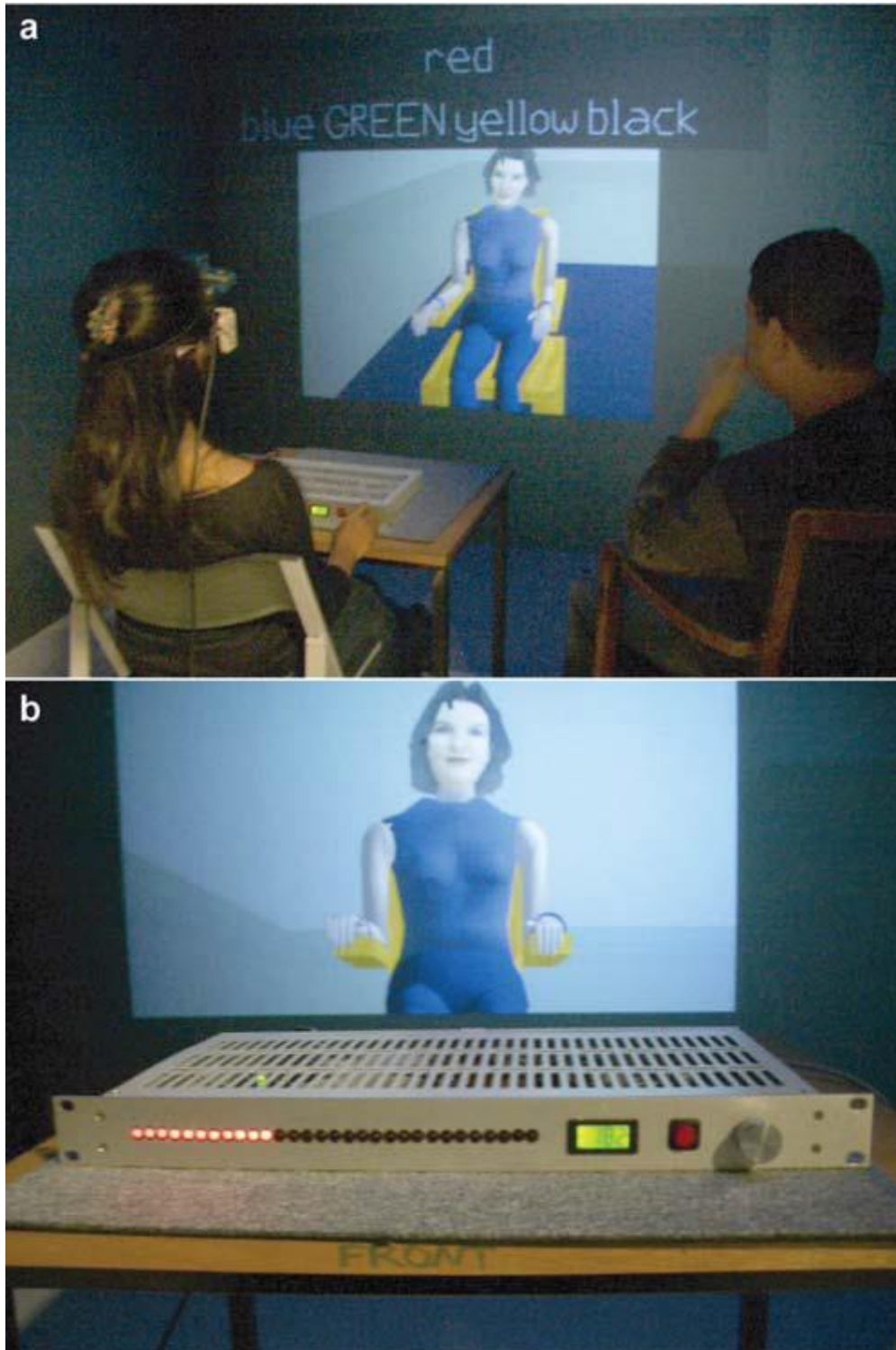


Figure 18. Scénario expérimental : (i) le participant est installé derrière une table comportant un générateur de choc électrique et l'expérimentateur est installé à droite du participant (a), (ii) l'élève virtuel apparaît en face du participant, (iii) le mot cible et la liste des quatre mots apparaissent au-dessus de l'élève virtuel (la bonne réponse est affichée en majuscule), (iv) après que le participant ait énoncé les cinq mots, l'élève virtuel donne sa réponse, (v) en cas de mauvaise réponse, le participant doit allumer la diode correspondant au voltage à administrer sur le générateur de choc (b), puis appuyer sur un bouton pour délivrer le choc à l'élève virtuel. (Slater et al., 2006)

En ce qui concerne le comportement d'obéissance, les résultats suggèrent un taux d'obéissance inférieur dans la condition « victime visible » (vs invisible). Les résultats montrent également une différence significative entre les deux conditions dans la volonté de mettre un terme à l'expérience, les participants de la condition « victime visible » ayant plus souvent déclaré à la fin de l'expérience avoir voulu mettre un terme à leur participation durant la procédure.

Relativement aux réponses physiologiques des participants, les auteurs se sont intéressés à deux aspects de l'activité électrodermale : (i) le niveau de conductance électrodermale (reflet du niveau général d'activation sympathique) et (ii) les réponses électrodermales (reflet des activations sympathiques transitoires ; e.g., Critchley, Elliott, Mathias, & Dolan, 2000 ; Nagai, Critchley, Featherstone, Trimble, & Dolan, 2004). Les résultats montrent que l'activité électrodermale des participants est significativement plus élevée sur ces deux aspects dans la condition « victime visible ». Ces résultats indiquent un plus fort niveau d'activation physiologique générale et de plus fortes réponses physiologiques après l'envoi des chocs électriques dans la condition « victime visible ». Afin d'obtenir des données sur la valence associée, les auteurs se sont également intéressés à deux aspects de l'activité cardiaque : (i) la fréquence cardiaque et (ii) la variabilité cardiaque. Les résultats montrent une hausse significative de la fréquence cardiaque dans la condition « victime visible », associée à une baisse significative de la variabilité cardiaque.

Ces résultats suggèrent que la réalité de la douleur observée est secondaire, ce qui est cohérent avec les études IRMf montrant que l'observation de signaux visuels indiquant l'expérience de la douleur par autrui active automatiquement les structures neurales responsables de la génération de la douleur (e.g., Lamm et al., 2011 ; Singer et al., 2004). Ainsi, il est raisonnable de faire l'hypothèse que l'observation de la douleur de l'avatar a suffi

à activer la neuromatrice de la douleur. Selon Slater et al. (2006), ces résultats « rouvrent la porte aux études empiriques directes de l'obéissance et des comportements sociaux extrêmes associés, un domaine de recherche inaccessible à l'étude expérimentale pour des raisons éthiques, *via* l'utilisation d'environnements virtuels ».

« *Immersive Video Milgram Obedience Experiment* »

Plus récemment, Dambrun et Vatiné (2010) ont élaboré un environnement vidéo immersif similaire à la condition 5 de Milgram (« *Immersive Video Milgram Obedience Experiment* » : IVMOE). La victime est ici non plus un avatar mais un acteur ayant été filmé jouant le rôle de l'élève. L'IVMOE est structuré en trois parties : (i) explications et instructions, (ii) tâche d'apprentissage, (iii) tâche expérimentale.

Au début de l'expérience, tous les participants sont *de facto* assignés au rôle de professeur, et une première interface vidéo les informe du fait que les chocs électriques et les réactions de l'élève (un acteur) seront simulés. Durant la première partie, Dambrun et Vatiné (2010) ont utilisé les mêmes explications et instructions que Milgram. Une interface présente la procédure expérimentale (*i.e.*, tâche de mémorisation et test de rappel) et les instructions (*i.e.*, appuyer sur le bouton rouge en cas d'erreur de l'élève, appuyer sur le bouton bleu en cas de bonne réponse). La dernière interface de cette première partie présente la configuration de l'interface expérimentale (*cf.* figure 19, sans vidéo). La deuxième partie commence avec l'installation de l'élève sur une chaise par un assistant de l'expérimentateur (*cf.* figure 19, Vidéo A). L'élève est attaché à la chaise et une électrode est fixée sur chacun de ses bras. La phase de mémorisation commence après l'installation, et les participants doivent regarder l'élève tenter de mémoriser trente paires de mots, paires énoncées séquentiellement par l'assistant de l'expérimentateur (*cf.* figure 19, Vidéo B). La troisième et dernière partie consiste à administrer un choc électrique à l'élève après chaque erreur, en augmentant le

voltage de 15 volts à chaque nouvelle erreur. Le « générateur de chocs » apparaît au bas de l'écran, reprenant à la fois les voltages et les étiquettes utilisés par Milgram (cf. figure 19).

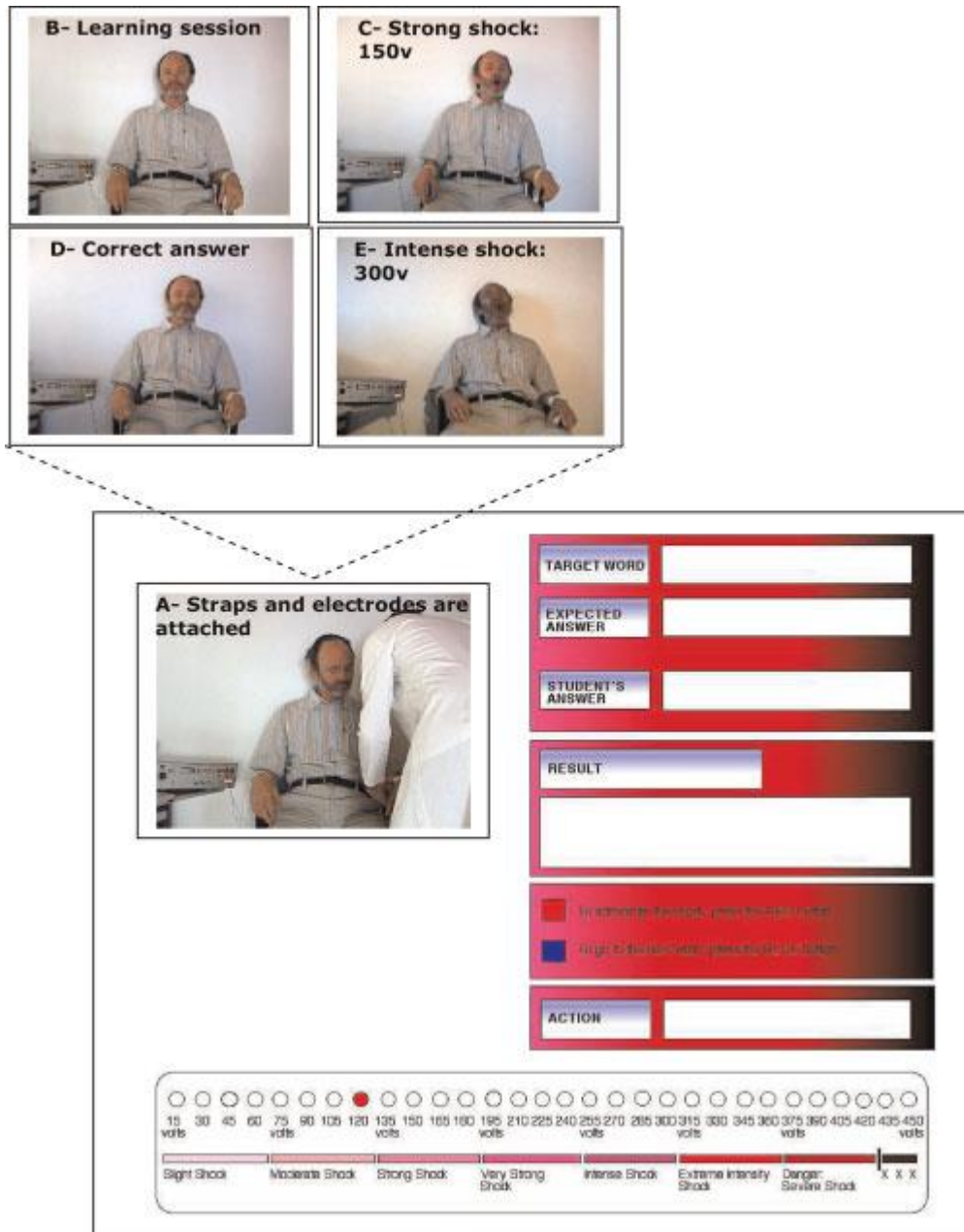


Figure 19. Interface de l'« Immersive Video Milgram Obedience Experiment » (IVMOE). (Dambrun & Vatiné, 2010)

Pour chaque paire de mots, les actions s'enchaînent de la manière suivante : (i) l'élève apparaît à l'écran et l'assistant de l'expérimentateur prononce le mot cible, (ii) 500ms après, le mot cible apparaît à l'écran suivi de la bonne réponse attendue, (iii) l'élève donne sa réponse, qui apparaît également à l'écran (500ms après), (iv) l'assistant de l'expérimentateur annonce la qualité de la réponse (« bonne réponse » ou « mauvaise réponse »), retour apparaissant à l'écran (fenêtre « Résultat »), (v) l'instruction d'appuyer sur le bouton rouge ou le bouton bleu apparaît à l'écran. En cas de bonne réponse, le professeur appuie sur le bouton bleu et l'essai suivant commence. En cas de mauvaise réponse, le professeur appuie sur le bouton rouge, un bruit de « buzz électrique » se fait entendre (durée de 1000ms) suivi d'un bruit de décharge électrique. Le professeur voit alors l'élève réagir en recevant le choc.

Les réponses de l'élève ainsi que ses réactions aux chocs suivent la procédure définie par Milgram : (i) l'élève donne une bonne réponse pour trois mauvaises réponses, (ii) à 150 volts, l'élève réagit fortement et demande à ce que l'expérience cesse (*cf.* figure 19, Vidéo C), (iii) quand l'élève donne une bonne réponse, il montre une certaine satisfaction (*cf.* figure 19, Vidéo D), (iv) à 300 volts, l'élève pousse un intense cri de douleur et refuse de répondre aux essais suivants (*cf.* figure 19, Vidéo E), (v) après 330 volts, il ne réagit plus et semble être dans un état d'agonie. Dans l'étude de Dambrun et Vatiné (2010), l'expérimentateur répondait au refus de poursuivre des participants par une séquence de deux injonctions (et non de quatre injonctions comme dans le paradigme de Milgram). La principale mesure était le choc maximum administré, de 0 (*i.e.*, refus d'administrer le moindre choc) à 30 (*i.e.*, 450 volts, pour les participants ayant administré la totalité des chocs).

Le principal objectif de Dambrun et Vatiné (2010) était de tester la validité et la fiabilité de l'IVMOE comme outil pour l'étude expérimentale de l'OD. Dans cette optique, les auteurs ont tenté de reproduire quelques résultats classiques en manipulant entre autres la

visibilité de la victime. Tout d'abord, les résultats montrent que le taux d'obéissance obtenu dans la condition « victime non visible » est similaire à celui obtenu par Milgram dans la condition « feedback vocal » (respectivement 53% et 62.5%). Conformément à ce qui a été observé dans des environnements réels, Dambrun et Vatiné (2010) ont trouvé que (i) la visibilité de la victime diminue significativement le taux d'obéissance (*cf.* tableau 10), (ii) le RWA prédit le niveau d'obéissance, (iii) les participants les plus obéissants attribuent plus de responsabilité à l'expérimentateur et à la victime, et moins de responsabilité à eux-mêmes. Les auteurs ont également trouvé que le niveau d'anxiété auto-reportée après l'expérience était significativement supérieur au niveau d'anxiété auto-reportée avant l'expérience.

Partant de ces résultats, Dambrun et Vatiné (2010) ont proposé que les participants obéissants dans un environnement immersif présentent les mêmes dispositions psychologiques que ceux ayant obéi dans le paradigme original de Milgram, et tendent à rationaliser leur comportement de la même manière. Comme le soulignent les auteurs, l'effet de « présence » provoqué par l'IVMOE est cohérent avec les travaux en neurosciences sociales de l'empathie montrant que l'observation des émotions d'autrui active automatiquement chez l'observateur les circuits neuronaux responsables de la génération de ces émotions. Il est raisonnable de faire l'hypothèse que ce mécanisme de résonance empathique aurait conduit les participants à faire l'expérience personnelle de la détresse de la victime et à éprouver de ce fait un « conflit psychologique » lors de l'envoi des chocs électriques. Dans leur ensemble, ces résultats suggèrent que l'IVMOE est un outil particulièrement adapté à l'étude expérimentale de l'OD.

Tableau 11
Résultats de l'étude de Dambrun et Vatiné (2010)

Niveau de choc	Mention verbale et niveau de voltage	Victime non visible (n = 40)	Victime visible (n = 40)
	Refus	2	
	Choc léger		
1	15	1	
2	30		
3	45		
4	60		1
	Choc modéré		
5	75	1	1
6	90		1
7	105		1
8	120		
	Choc fort		
9	135		2
10	150		3
11	165	1	
12	180	1	1
	Choc très fort		
13	195		
14	210		
15	225		1
16	240		
	Choc intense		
17	255		
18	270		
19	285		
20	300		1
	Choc extrêmement intense		
21	315		
22	330		
23	345		1
24	360	1	
	Attention : choc dangereux		
25	375		
26	390		
27	405		1
28	420		
	XXX		
29	435		
30	450	8	2
Moyenne des niveaux de chocs maxima :		19.53	14.19
Pourcentage d'obéissance :		53%	13%

PARTIE 2. PROBLEMATIQUE GENERALE

Cheetham et al. (2009) ont proposé que l'obéissance dans le paradigme de Milgram puisse être (en partie) la conséquence d'un déficit de régulation de la détresse induite par l'observation de la douleur de la victime. Nous avons vu dans l'introduction de ce travail doctoral que les capacités d'autorégulation déterminent les réponses affectives-motivationnelles à la douleur d'autrui : (i) un niveau moyen d'activation sympathique permet l'initiation d'une réponse orientée vers autrui (sympathie), (ii) un état d'hyperactivation sympathique génère des réponses émotionnelles aversives centrées sur soi (détresse personnelle ; Decety & Lamm, 2009 ; Eisenberg & Eggum, 2009). Des travaux suggèrent que la détresse personnelle affecte le comportement *via* le recours à des stratégies d'évitement (e.g., Eisenberg et al., 2004, 2006). Les stratégies d'autorégulation basées sur l'évitement sont une réponse courante et maladaptative aux expériences émotionnelles aversives (e.g., Aldao et al., 2015 ; Barlow, 2002 ; Gratz & Roemer, 2004). Ces stratégies incluent la distraction, l'inhibition des réponses internes, la suppression des expressions faciales émotionnelles et les comportements d'évitement (e.g. Gamez et al., 2011; Hayes et al., 1999). Des études en imagerie cérébrale ont montré que l'observation de la douleur d'autrui dans une situation « sans issue » induit un phénomène de « désactivation » de l'AMY, phénomène reflétant une stratégie cognitive d'adaptation à des événements douloureux mais inévitables (e.g., Petrovic et al., 2004). Or, un moindre engagement de l'AMY dans une situation mettant en jeu le respect de l'intégrité physique d'autrui a été associé à une moindre aversion pour l'atteinte à autrui (e.g., Blair, 2007 ; Mendez, 2006). Une telle stratégie d'autorégulation dans le cadre du paradigme de Milgram pourrait ainsi faciliter l'OD.

Les travaux de Decety et ses collaborateurs ont contribué à mettre en évidence la diversité des structures cérébrales et des mécanismes affectifs et cognitifs impliqués dans l'empathie (e.g., Decety & Lamm, 2009). Le modèle de Decety et Jackson (2004) insiste entre autres sur l'importance de la flexibilité cognitive et de l'autorégulation pour la gestion et

l'optimisation du partage de la douleur. Or, l'autorégulation dépendrait de manière critique de la flexibilité physiologique (e.g., Gross, 1998). La théorie polyvagale (e.g., Porges, 2007) insiste sur le rôle du nerf vague dans la stabilisation des processus autonomes et la libération des ressources nécessaires aux tendances prosociales (sentiments prosociaux comme la sympathie, comportement d'approche). Le modèle d'intégration neuroviscérale (e.g., Thayer & Lane, 2000) propose que l'influence inhibitrice du CPF sur les circuits sympatho-excitateurs module la variabilité de la fréquence cardiaque *via* la branche myélinisée du nerf vague. L'inhibition des circuits sympatho-excitateurs est essentielle pour permettre à l'organisme de se détacher des stratégies par défaut orientées vers la menace et ainsi pour l'expression de tendances prosociales. De manière cohérente avec ces deux approches, de nombreuses études empiriques ont mis en évidence qu'un moindre tonus vagal favorise l'anxiété sociale, le stress psychosocial, la détresse personnelle en réponse à autrui en détresse et l'adoption de stratégies d'autorégulation basées sur l'évitement en réponses à des stimuli aversifs (e.g., Aldao et al., 2015 ; Fabes & Eisenberg, 1997 ; Fabes et al., 1993 ; Movius & Allen, 2005 ; Thayer & Lane, 2007).

L'introduction de ce travail doctoral nous a également permis de mettre en évidence les manques et insuffisances d'une recherche sur l'OD confrontée à un certain nombre d'obstacles éthiques et méthodologiques. Dans cette perspective, la technologie d'environnement immersif représente un outil particulièrement adapté à l'étude expérimentale de l'OD (e.g., Dambrun & Vatiné, 2010 ; Slater et al., 2006).

La question centrale de cette thèse porte sur le rôle des mécanismes d'autorégulation dans une situation d'OD. Pour comprendre celui-ci, nous avons interrogé le rôle des variations dans la vulnérabilité au stress dans l'émergence des différences interindividuelles face à des ordres destructeurs. Les différences interindividuelles dans la flexibilité

physiologiques ont été retenues comme facteur explicatif et prédictif de la vulnérabilité au stress. Notre approche peut être qualifiée de biopsychosociale : elle propose en effet que la flexibilité physiologique (aspects biologiques) participe aux réactions à la douleur d'autrui (aspects psychologiques) qui à leur tour guident les réponses comportementales dans le paradigme de Milgram (aspects sociaux).

Ainsi, un premier questionnaire porte sur le *tonus vagal*. Si l'OD est déterminée par la vulnérabilité au stress des individus, le tonus vagal des participants devrait prédire leur comportement d'OD. Plus précisément, un moindre tonus vagal devrait être associé à un taux supérieur d'OD. Le RWA étant un prédicteur de l'OD, un moindre tonus vagal devrait également être associé à des scores supérieurs de RWA.

Un deuxième questionnaire porte sur le *contrôle inhibiteur*. Si les participants obéissants ont un moindre tonus vagal, ils devraient avoir recours au contrôle inhibiteur afin de diminuer leur inconfort. Or, cette stratégie représente un effort cognitif. La recherche ayant montré que l'effort cognitif augmente la variabilité cardiaque dans le domaine temporel (e.g., Kalinichenko & Yur'eva, 2012 ; Porges, 2007 ; Segerstrom & Ness, 2007), nous devrions retrouver une telle augmentation du tonus vagal chez les participants obéissants au cours de la procédure d'obéissance. Si cette stratégie permet l'inhibition des réponses internes, nous devrions également observer une diminution de la fréquence cardiaque chez les participants obéissants. Nous pouvons aussi nous attendre à ce que le contrôle inhibiteur médiate l'influence du tonus vagal et du RWA sur l'OD. Enfin, la manipulation expérimentale de la capacité des participants à fournir un tel effort au contrôle inhibiteur devrait nous permettre d'influencer l'OD. Plus précisément, réduire l'exercice du contrôle inhibiteur *via* l'induction d'une « fatigue mentale » (e.g., Francis & Inzlicht, 2016) devrait favoriser la désobéissance.

Un dernier questionnement porte sur les *mécanismes neuronaux* sous-tendant l'OD. Si la mesure des processus physiologiques périphériques peut contribuer à la compréhension des mécanismes responsables de l'OD, la mesure des processus neurophysiologiques engagés durant la procédure d'obéissance peut permettre une mesure plus directe du contrôle inhibiteur et de l'*aversion pour l'atteinte à autrui*. Les ondes thêta étant impliquées dans l'inhibition des réponses émotionnelles aversives (e.g., Ertl et al., 2013), leur amplitude devrait être plus importante chez les participants obéissants. Enfin, l'activité hémodynamique au niveau des aires cérébrales impliquées dans l'anticipation et l'évaluation des conséquences des décisions morales pour autrui devrait également permettre de distinguer participants obéissants et désobéissants. Plus précisément, nous pouvons nous attendre à ce que l'activité hémodynamique au niveau du CPFVM (et plus particulièrement au niveau du COF) soit moindre chez les participants obéissants.

Afin d'apporter des réponses à ces différents questionnements, nous avons réalisé six études : (i) une étude sur la relation entre autoritarisme et utilitarisme moral (étude 1), et une autre sur la relation entre autoritarisme et tonus vagal (étude 2), (ii) une étude sur la relation entre tonus vagal, réactivité vagale et OD (étude 3), (iii) une étude sur l'effet d'un affaiblissement temporaire du contrôle inhibiteur (*via* l'induction expérimentale d'une fatigue mentale) sur l'OD (étude 4), (iv) une étude EEG sur la relation entre amplitude des ondes thêta et OD (étude 5), (v) une étude par spectroscopie proche infrarouge de la relation entre OD et réponse hémodynamique au niveau du CPFDL, du CPFVM et du COF (étude 6). Nos études sur la soumission à l'autorité reposent sur l'usage d'une version alternative de l'IVMOE. Le choix d'une approche guidée aux niveaux théorique et méthodologique par les neurosciences sociales et le recours à l'environnement immersif pour l'étude expérimentale de l'OD visent à maximiser la validité et la fiabilité de ce travail doctoral.

Les parties 3-5 de ce document de thèse ont été rédigées sous la forme d'articles scientifiques. Chaque partie peut donc se lire de manière indépendante, ce qui peut expliquer un recouvrement partiel entre les différentes parties.

Les autorisations des différents comités d'éthique sollicités se trouvent en annexe de ce document (*cf.* annexes 1 et 2). L'environnement vidéo immersif utilisé pour les études 3-6 est présenté dans les manuscrits des études correspondantes. Nous avons inséré en annexe la « notice d'information » remise aux participants de ces quatre études, dont la fonction était de justifier la procédure d'obéissance et de préciser les consignes à respecter (*cf.* annexe 3). Les modalités et outils de recrutement des participants sont également présentés en annexe (*cf.* annexe 4).

PARTIE 3. AUTORITARISME, UTILITARISME MORAL ET TONUS VAGAL

En amont de nos études sur la soumission à l'autorité, il nous a semblé important d'accéder à une meilleure compréhension des aspects émotionnels de l'autoritarisme en tant que prédicteur de l'OD. Nous avons vu dans la partie introductive de ce travail doctoral que RWA et SDO prédisent des phénomènes sociopolitiques similaires tout en constituant des attitudes distinctes (e.g., Duckitt & Sibley, 2010). Le RWA serait l'expression dans la sphère sociale d'une vulnérabilité au stress (e.g., Jugert & Duckitt, 2009) alors que la SDO serait associée à un défaut de résonance empathique (e.g., Chiao et al., 2009). Or, nous avons également vu que le comportement prosocial dépend de la capacité des individus à éprouver un sentiment d'aversion pour l'atteinte à autrui et à ajuster leur niveau d'excitation physiologique à un instant précis (flexibilité autonome). Nous avons ainsi réalisé une première étude portant sur la relation entre autoritarisme et utilitarisme moral (étude 1), et une deuxième étude explorant le lien entre autoritarisme et tonus vagal (étude 2).

3.1. Autoritarisme et Utilitarisme moral

Objectifs et présentation de l'étude

L'objectif de cette étude était d'explorer la relation entre autoritarisme et utilitarisme moral. L'utilitarisme moral ayant été conceptualisé comme la conséquence d'une moindre aversion pour l'atteinte à autrui (e.g., Moll & de Oliveira-Souza, 2007), nous avons fait l'hypothèse que seul la SDO devrait être reliée à l'utilitarisme moral. Nous nous attendions également à ce qu'un moindre altruisme dispositionnel médiate la relation entre SDO et utilitarisme. Afin de tester ces hypothèses, les participants ($n = 500$) ont participé à une étude en ligne. Dans un premier temps, ces derniers devaient répondre à un questionnaire incluant les versions françaises des échelles SDO et RWA ainsi qu'une mesure d'altruisme dispositionnel. Dans un second temps, ils devaient répondre à deux versions du « dilemme du tramway » : (i) la version « aiguillage » impliquant une atteinte à autrui indirecte et

« impersonnelle », (ii) la version « pont » impliquant une atteinte directe et « personnelle ».

Suivant la procédure préconisée par Tassy, Deruelle, Mancini, Leistedt et Wicker (2013), nous avons dissocié deux aspects de l'évaluation morale : (i) le jugement évaluatif (*i.e.*, « Est-il acceptable de... ? ») impliquant une évaluation objective à un niveau abstrait (le participant adopte un référentiel allocentré), (ii) le choix d'action (*i.e.*, « Accepteriez-vous de... ? ») impliquant une évaluation subjective (le participant adopte un référentiel égocentré et ainsi un mode de résolution subjectif). L'importance d'une telle dissociation a été mise en lumière par des études montrant que la psychopathie est associée à une capacité de jugement évaluatif préservée mais des choix comportementaux inadaptés du point de vue moral (voir Tassy et al., 2013). La recherche ayant montré que l'anticipation des conséquences socio-émotionnelles d'une action est cruciale pour la décision morale dans les situations de dilemme (e.g., Koenigs et al., 2007), nous avons ajouté une mesure de sentiment de culpabilité en cas de réalisation de l'action proposée pour chaque dilemme.

Manuscrit 1⁶

⁶ Manuscrit soumis pour publication auprès de la revue *Social Justice Research*.

Ideology, Altruism and Utilitarian Moral Reasoning

Johan Lepage¹, Laurent Bègue¹, Michael Dambrun², & Martial Mermillod^{3,4}

¹ Laboratoire InterUniversitaire de Psychologie, Université Grenoble Alpes, Grenoble, France

² Laboratoire de Psychologie Sociale et Cognitive, CNRS UMR 6024, Université Blaise Pascal, Clermont-Ferrand, France

³ Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105, Université Grenoble Alpes, Grenoble, France

⁴ Institut Universitaire de France, Paris, France

Corresponding author

Johan Lepage, PhD Candidate: Laboratoire InterUniversitaire de Psychologie, UFR SHS, 1251 avenue Centrale, BP 47, 38040 Grenoble Cedex 9, France; E-mail: johan.lepage@gmail.com, Phone: (+33 4) 76 82 58 92, Fax: (+33 4) 76 82 56 65

Abstract

The present study aimed to examine (i) the association between ideology and moral decision-making in Trolley-type dilemmas, (ii) the potential mediating role of dispositional altruism. Participants ($n = 500$) first completed a survey assessing social dominance orientation (SDO) and right-wing authoritarianism (RWA), guilt and dispositional altruism. Our experimental design dissociated two moral scenarios (impersonal dilemma and personal dilemma) as well as two aspects of moral evaluation (“evaluative judgment” and “choice of hypothetical action”). We found that high SDO individuals favored the utilitarian options in both personal and impersonal dilemmas and reported lower anticipation of the social-emotional consequences of the proposed action (i.e., lower guilt). Dispositional altruism fully mediated the association between SDO and utilitarian responses. All in all, these findings provide a coherent harm aversion account of the association between ideology and moral decision.

Keywords

Social dominance orientation, right-wing authoritarianism, moral utilitarianism, altruism, guilt, harm aversion

Introduction

The political division between liberals and conservatives is ancient and universal (e.g., Bobbio, 1996) and a great deal of research has been devoted to explaining the causes of this antagonism. Empirical research demonstrates that liberals and conservatives differ on many physiological and psychological aspects (e.g., Hibbing, Smith, & Alford, 2013; Jost & Amodio, 2012; Jost, Glaser, Kruglanski, & Sulloway, 2003; Schwartz, Caprara, & Vecchione 2010). Recent research suggests that differences in moral cognition could be responsible for many of the differences between liberals and conservatives. For instance, Graham, Haidt and Nosek (2009) hypothesized and found that individuals' position on the left–right political spectrum is associated with different moral domains that are considered most relevant in morality: liberals are primarily concerned with “Harm/care” and “Fairness/reciprocity”, whereas conservatives also consider the domains of “Ingroup/loyalty”, “Authority/respect” and “Purity/sanctity”, as highly relevant for moral judgment. Unfortunately, potential differences in other dimensions of moral functioning remain underexplored.

Dualistic approach of moral decision

Individuals daily face social situations involving conflicts between two major traditions in normative ethics: (i) *utilitarianism* evaluates actions based on their outcomes (utilitarian decisions aim to achieve the greatest amount of happiness for the most amount of people, regardless of the severity of the actions to achieve such a goal), (ii) *deontology* evaluates the actions themselves (certain actions are absolutely wrong, regardless of the consequences; Crockett, 2013). Moral dilemmas used in research are characterized by the presence of a conflict “in which either decision would result in the transgression of an ethical imperative, often resulting in harm to others” (Tost & Meyer-Lindenberg, 2010). One prominent case of a moral dilemma involves a speeding trolley down a track, on route to run

over five persons. In the switch version (“impersonal moral dilemma”), the only way to save these five persons is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person. In the footbridge version (“personal moral dilemma”), the only way to save the five people is to push a stranger off a bridge, onto the tracks below. The stranger will die but his body will stop the trolley from killing the five. While most of people presented with the switch problem endorse pressing the switch, this pattern is reversed for the footbridge problem (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001).

Brain imaging studies have shown that personal moral dilemmas produced increased activity in brain areas associated with emotional processing, while impersonal moral dilemmas induced increased activity in brain areas associated with executive functions (e.g., Greene, 2009). Other studies found that impairment of several areas of the “affective moral brain” induced increased utilitarianism (e.g., ventromedial prefrontal cortex, amygdala; Blair, 2007; Funk & Gazzaniga, 2009; Moll & de Oliveira-Souza, 2007). These results are used to support the dominant dualistic approach to moral decision which holds that automatic emotional processes pull for deontological decisions while controlled rational/cognitive processes pull for utilitarian responses (e.g., Greene, 2009).

Utilitarianism is believed to require increased deliberative reflective reasoning (Paxton, Ungar, & Greene, 2012) and to be driven by a general concern with maximizing the good. It has been argued that empirical evidence about the psychological underpinnings of moral decision-making should lead us to adopt an utilitarian approach to ethics (Greene, 2008; Singer, 2005). However, recent research casts doubt on the assumption that utilitarianism is driven by an impartial concern for the greater good, and provides some evidence that the ability to feel concern for the other’s suffering is likely a stronger predictor of utilitarian decisions than the integrity of one’s rational/cognitive system. Studies found that

utilitarian decision is associated with increased egoist concern (e.g., Kahane, Everett, Earp, Farias, & Savulescu, 2015), antisocial personality traits such as psychopathy and Machiavellianism (e.g., Bartels & Pizarro, 2011), as well as with reduced empathic concern (an other-oriented motivated response to the others' pain and distress), higher levels of testosterone and lower levels of serotonin (Carney & Mason, 2010; Crockett, Clark, Hauser, & Robbins, 2010; Montoya et al., 2013).

Whether utilitarian responding is more strongly linked to increased deliberative reasoning or decreased aversion to harming others remains an open question (e.g., Duke & Bègue, 2015). The present study aimed to explore (i) the association between ideology and utilitarianism, (ii) the potential mediating role of dispositional altruism.

The core dimensions of ideology

Political psychology research indicates that the left–right dimension is an oversimplification of the spectrum of political thought and that ideology is better captured by two dimensions of ideology: right-wing authoritarianism (RWA) and social dominance orientation (SDO; e.g., Duckitt, 2001). RWA and SDO both predict a wide range of sociopolitical phenomena such as prejudice, right-wing political party preference, patriotism, militarism, or support for capital punishment (e.g., Altemeyer, 1998; Duckitt, 2001). However, the Dual Process Motivational model (e.g., Duckitt & Sibley, 2010) posits that RWA and SDO are two distinct ideological attitudes with each expressing motivational goals and values made chronically salient for individuals by their worldviews and personality traits. High RWA expresses the motivational goals and values of establishing or maintaining collective security, which are made chronically salient by the belief that the world is inherently dangerous and threatening. The predisposing personality dimension is social conformity (i.e., low Openness to experience and high Conscientiousness). In contrast, high SDO expresses the motivational

goals and values of power, dominance, and superiority over others, chronically salient because of a view of the world as a ruthlessly competitive jungle in which the strong win and the weak lose. SDO stems from the underlying personality dimension of tough vs. tender mindedness (i.e., low Agreeableness). Studies also found that SDO is associated with antisocial personality traits such as Psychopathy and Machiavellianism (e.g., Hodson, Hogg, & MacInnis, 2009).

SDO seems to qualify as a potential predictor of unethical decision and aggressive behavior. The capacity of resonating with the others' suffering provides the affective-motivational base for care-based moral reasoning (e.g., Blair, Mitchell, & Blair, 2005; Eisenberg & Morris, 2001; Hoffman, 2000) and studies using subjective self-report suggest that SDO is negatively related to concern for the others' well-being (e.g., Pratto, Sidanius, Stallworth, & Malle, 1994; Sidanius & Pratto, 1999). An fMRI study found an inverse relation between SDO scores and neural responses when perceiving pain in others within fronto-insular regions critical to the ability to share and feel concern for the other's suffering (anterior cingulate cortex and anterior insula; Chiao, Mathur, Harada, & Lipke, 2009). Others studies have shown that SDO bolstered support for war by reducing concern for its human costs (e.g., McFarland, 2005). Few more recent studies investigated the association between authoritarianism and moral intuitions (Haidt & Graham, 2007) and found that SDO was negatively related to concerns about fairness and avoidance of harm (Federico, Weber, Ergun, & Hunt, 2013; Milojev et al., 2014; see also Kugler, Jost & Noorbaloochi, 2014).

Some studies suggest that RWA is also related to behavioral aggression. For instance, RWA scores correlated with severity of sentences recommended for lawbreakers (but the correlation was not found if the lawbreakers were government officials; Altemeyer, 1981, 1988). RWA also correlated with selecting more severe shock levels to punish a learner for mistakes on

a task (Altemeyer, 1981, 1988). However, the interpretation that these studies support a direct link between RWA and aggression seems questionable: (i) the sentences level advocated was typically low and seem “more parsimoniously interpreted as reflecting greater favorability to the authorities and lower favorability to persons challenging or violating authority” (Duckitt, 2001); (ii) the shocks level selected was also weak, and “may simply reflect an association between RWA and a belief in the efficacy of punishment in learning situations rather than indicating a trait of authoritarian aggression” (Duckitt, 2001); (iii) Altemeyer (1988) found that general levels of aggression irrespective of target were uncorrelated with RWA scores. Passini (1988) examined the association between SDO and the multidimensional structure (conventionalism, authoritarian submission, authoritarian aggression) of Altemeyer’s (1981, 1988) conceptualization of RWA, and found that authoritarian aggression was related to SDO. According to the author, this result is a clarification of the classical correlation between RWA and SDO, and suggests that aggression against deviants is a possible consequence of RWA and not one of its components.

All in all, previous research suggests that low aversion to personally harming others favors utilitarian responding, and that SDO is related to low concern for the others’ well-being. Thus, one could hypothesize that high SDO individuals may favor utilitarian options in moral dilemmas pitting utilitarian outcomes (saving five lives) against highly aversive harmful actions (killing an innocent person) because of low altruistic concern.

Altruism

The empathy-altruism hypothesis (e.g., Batson, 1991) states that empathic concern results in an altruistic motivation (i.e., a motivational state in which the goal is to increase another person’s welfare as an end in itself). This hypothesis is well supported by a large body of research, demonstrating that sympathetic responses to the others’ suffering lead to

altruistically motivated helping responses (e.g., Eisenberg & Eggum, 2009). However, many researchers disagree with Batson's arguments regarding the existence of true altruism (e.g., Maner et al., 2002) and the dominant view is that any seemingly altruistic behavior must have been driven by some kind of selfish motivation (e.g., individuals adopt seemingly altruistic behavior to alleviate one's own but not the other's distress). This view is recognizable in an influential school of biology founded by Huxley (1894) which states that humans are born selfish and that morality relies on cognitive and cultural processes (e.g., Høgh-Olesen, 2010). By contrast, a Darwinian approach posits "that morality is a natural outgrowth of the social instincts, hence continuous with the sociality of other animals" (de Waal, 2003). Consistently, comparative and neuroscience research "demonstrates that many of our highest moral, altruistic and sympathetic inclinations may be evolutionary anchored in mammalian sociality and based on pre-programmed, emotional and neuro-endocrine feedbacks – operating as spontaneous "gut-feelings" – rather than pure reason, and just as selected for as our more selfish impulses" (Høgh-Olesen, 2010).

Altruism has also been examined as a trait-like variable that predisposes an individual to respond behaviorally with a predictable pattern. Within the HEXACO model of personality (e.g., Ashton & Lee, 2007), altruism refers to a tendency to be sympathetic, to avoid causing harm and to react with generosity toward the others in pain or distress, and to accept personal drawbacks for the sake of others. Empirical studies show that measures of dispositional altruism predict prosocial behavior in the laboratory as well as in the field (e.g., Staub, 1974; Penner & Finkelstein, 1998). Studies also found that altruism as conceptualized within the HEXACO model of personality is related to reciprocity, generosity (e.g., Baumert, Schlösser, & Schmitt, 2014) as well as to political orientation (more altruistic individuals placed themselves further left on the political spectrum; Zettler & Hilbig, 2010).

Hypotheses

The relationship between ideology and utilitarianism has rarely been investigated (for an exception, see Bostyn, Roets, & Van Hiel, 2016). *First*, we explored the association between authoritarian attitudes and moral responding to trolley-type dilemmas in a large sample using multiple regression analyses. Because high SDO individuals may perceive the potential harm of others as less impermissible, we hypothesized that higher SDO scores would predict increased utilitarianism. *Second*, we investigated the potential mediating role of dispositional altruism using the bootstrapping method advocated by Preacher and Hayes (2004). In line with the empathy-altruism hypothesis, we expected that low dispositional altruism would mediate the influence of SDO on utilitarian responding.

Material and method

Five hundred French citizens (293 female; $M_{age} = 26.65$, $SD = 5.41$) participated in an online study (Inquisit Web). RWA was assessed using 10 items ($\alpha = .83$) from the French translation of Altemeyer's 20-item (1988) RWA Scale (1 = *totally disagree* to 7 = *totally agree*). SDO was assessed using 10 items ($\alpha = .84$) from the French translation of the 16-items version (Duarte, Dambrun, & Guimond, 2004) of the SDO Scale (1 = *totally disagree* to 7 = *totally agree*). Altruism was assessed using the French translation (Boies, Yoo, Ebacher, Lee, & Ashton, 2004) of the 100 items ($\alpha = .82$) HEXACO-PI-R (1 = *totally disagree* to 5 = *totally agree*). Altruism is located interstitially among honesty-humility, agreeableness and emotionality. Therefore, "the overall tendency to be altruistic or to be antagonistic will represent a blend of those three dimensions" (Lee & Ashton, 2006).

Participants responded to two moral scenarios designed to contrast the impact of emotional salience on moral judgment: (i) emotionally salient "personal" harm that involved harming one person to benefit many ("Footbridge dilemma") and (ii) less emotionally salient

“impersonal” harm that involved harming one person to benefit many (“Trolley dilemma”). Following the procedure advocated by Tassy, Deruelle, Mancini, Leistedt and Wicker (2013), our experimental design dissociates two aspects of moral evaluation: (i) evaluative judgment (i.e., “Would it be *acceptable* to...?”) that involves an impersonal objective evaluation, (ii) a first-person perspective choice of hypothetical action (“Would *you do*...?”). Participants indicated their preferences by clicking on a box (1 = *No, absolutely not* to 5 = *Yes, absolutely*). The importance of dissociation between “evaluative judgment” and “choice of hypothetical action” was highlighted by studies showing that the association between high psychopathic traits and increased utilitarianism is restricted to choice (e.g., Tassy et al., 2013).

We have added a measure of guilt (i.e., “Would you feel guilty if you perform this action?”). This higher-order emotion seems clearly theoretically relevant to utilitarianism. Ferguson and Stegge (1998) defined guilt as “an agitation-based emotion or painful feeling of regret that is aroused when the actor actually causes, anticipates causing, or is associated with an aversive event”, and prior work found that the anticipation of the social-emotional consequences of the proposed action (guilt) is critical for reasoning about moral dilemmas (e.g., Koenigs al., 2007). Participants responded by clicking on a box (1 = *No, absolutely not* to 5 = *Yes, absolutely*). We expected that higher SDO would be related to lower guilt.

Results

Associations between ideological attitudes and responses to moral dilemmas

Descriptive statistics and Pearson’s correlations are presented in Table 1. Results are presented in Table 2. Table 1 shows that RWA and SDO were positively correlated with each other. The variance inflation factor (VIF) score was 1.26, indicating that multicollinearity was not a concern. Replicating previous findings, we found that the participants judged personal harms as less permissible than impersonal harms, $t(499) = -16.70, p < .001, 95\% \text{ CI } [-1.09, -$

.86]. They also provided fewer endorsement of personal harms than of impersonal harms, $t(499) = -17.22, p < .001, 95\% \text{ CI} [-1.19, -.94]$. Table 2 shows that higher SDO scores predict higher acceptability and endorsement of the utilitarian options in both impersonal and personal dilemmas.

Table 1

Descriptive statistics and correlations between authoritarian attitudes, dispositional altruism and moral outcomes

	1	2	3	4	5	6	7	8	9
1 RWA	—	.46***	-.04	.02	.02	-.10*	-.002	.01	-.05
2 SDO		—	-.33***	.09*	.11*	-.14***	.08	.09*	-.15***
3 Altruism			—	-.14***	-.11**	.04	-.16***	-.19***	.005
4 TJ				—	.74***	.06	.49***	.39***	.18***
5 TC					—	.06	.39***	.44***	.19***
6 TG						—	.05	.06	.63***
7 FJ							—	.77***	.07
8 FC								—	.10*
9 FG									—
<i>M (SD)</i>	2.88 (1.05)	2.31 (.99)	3.68 (.65)	2.96 (1.37)	2.92 (1.42)	3.98 (1.34)	1.98 (1.21)	1.85 (1.17)	4.06 (1.34)

Note. RWA = right-wing authoritarianism; SDO = social dominance orientation; TJ = trolley judgment; TC = trolley choice; TG = trolley guilt; FJ = footbridge judgment; FC = footbridge choice; FG = footbridge guilt. * $p < .05$ ** $p < .01$ *** $p < .001$

Table 2

Multiple regression examining the associations between ideological attitudes and responses to moral dilemmas

Associations between ideological attitudes and responses to the Trolley dilemma					
Criterion variables	Predictor variables	β	t	p	95% CI
Judgment	SDO	.10	2.12	.03	[.01, .28]
Choice		.12	2.56	.01	[.04, .32]
Guilt		-.12	-2.51	.01	[-.30, -.03]
Judgment	RWA	-.02	-.48	.62	[-.16, .10]
Choice		-.03	-.63	.52	[-.18, .09]
Guilt		-.04	-.97	.33	[-.19, .06]
Associations between ideological attitudes and responses to the Footbridge dilemma					
Criterion variables	Predictor variables	β	t	p	95% CI
Judgment	SDO	.10	2.06	.03	[.006, .24]
Choice		.10	2.13	.03	[.01, .24]
Guilt		-.16	-3.39	.001	[-.35, -.09]
Judgment	RWA	-.05	-.98	.32	[-.18, .06]
Choice		-.03	-.61	.53	[-.15, .08]
Guilt		.02	.55	.57	[-.09, .16]

Note. Continuous predictors were centered before analysis. SDO = social dominance orientation; RWA = right-wing authoritarianism.

Mediation analyses

Regression analyses resulted in a significant effect of SDO on judgment and choice in both moral dilemmas. We tested the extent to which altruism mediated the effect of SDO on moral decision-making.

Judgment (Trolley dilemma). In order to test and estimate path *a*, we used altruism as the criterion variable and SDO as predictor, and found that higher SDO predicted lower altruism, $b = -.33, t = -7.81, p < .001, 95\% \text{ CI } [-.41, -.24]$. In order to test and estimate path *b* and path *c'*, we used Judgment as the criterion variable, and SDO and altruism as predictors. The relationship between altruism and Judgment remained significant even while controlling for SDO, $b = -.17, t = -2.77, p = .005, 95\% \text{ CI } [-.30, -.05]$. Most importantly, the effect of SDO on Judgment controlling for altruism became non-significant, $b = .07, t = 1.11, p = .26, 95\% \text{ CI } [-.05, .19]$. Employing the bootstrapping method (with 5000 samples) produced a 95% bias-corrected confidence interval of [.01, .11] for the indirect effect of SDO on Judgment through altruism ($z = 2.59, p = .009$). The confidence interval excluded zero, which suggests a significant indirect effect (see Figure 3a).

Choice (Trolley dilemma). The bootstrapping method didn't provide evidence for an indirect effect of SDO on Choice through altruism (95% CI [-.001, .09], $z = 1.83, p = .06$).

Judgment (Footbridge). The relationship between altruism and Judgment remained significant even while controlling for SDO, $b = -.18, t = -3.20, p = .001, 95\% \text{ CI } [-.29, -.07]$. The effect of SDO on Judgment controlling for altruism became non-significant, $b = .03, t = .67, p = .50, 95\% \text{ CI } [-.07, .15]$. Employing the bootstrapping method (with 5000 samples) produced a 95% bias-corrected confidence interval of [.01, .11] for the indirect effect of SDO on Judgment through altruism ($z = 2.94, p = .003$) (see Figure 3b).

Choice/Footbridge. The relationship between altruism and Choice remained significant even while controlling for SDO, $b = -.21$, $t = -3.93$, $p < .001$, 95% CI [-.32, -.10]. The effect of SDO on Choice controlling for altruism became non-significant, $b = .03$, $t = .70$, $p = .48$, 95% CI [-.06, .14]. Employing the bootstrapping method (with 5000 samples) produced a 95% bias-corrected confidence interval of [.03, .12] for the indirect effect of SDO on Judgment through altruism ($z = 3.49$, $p < .001$) (see Figure 3c).

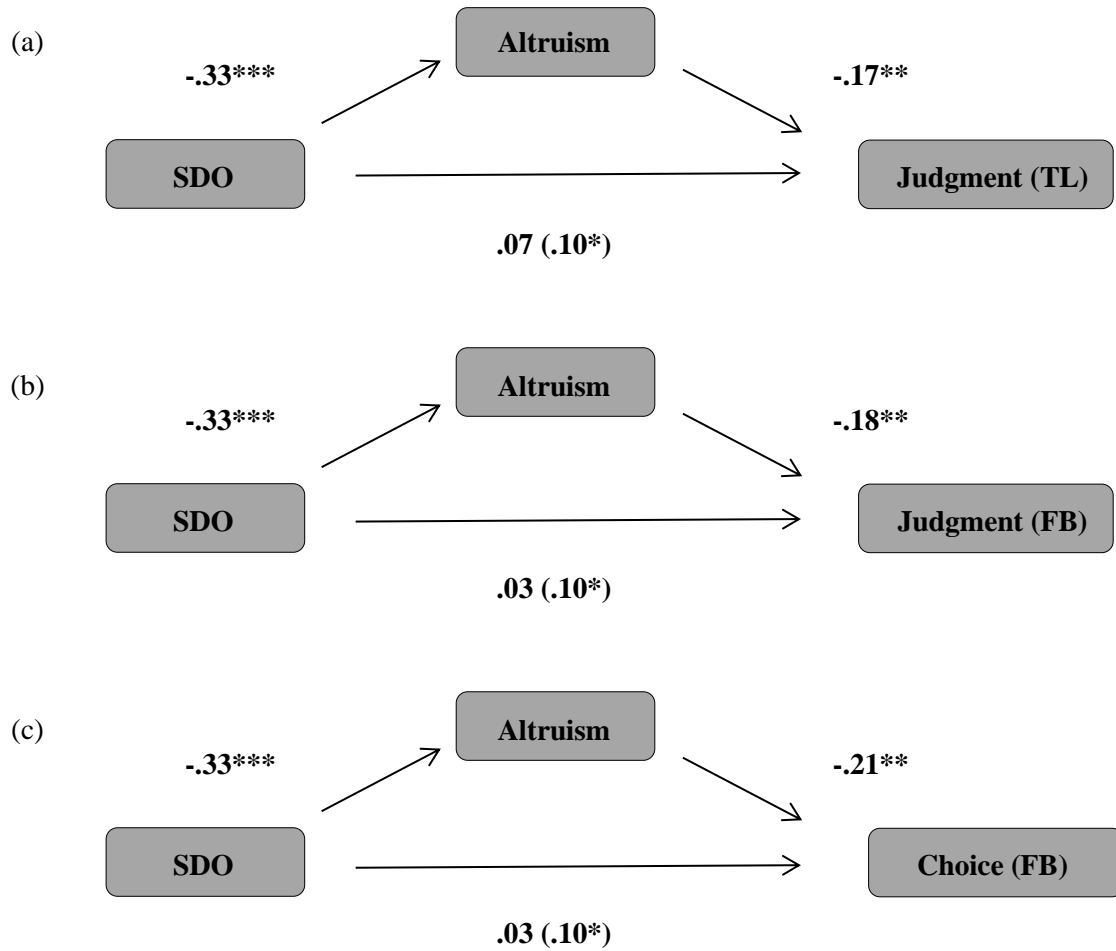


Figure 1. Standardized regression coefficients (a) for the relationship between SDO and Judgment (Trolley dilemma) as mediated by altruism, (b) for the relationship between SDO and Judgment (Footbridge dilemma) as mediated by altruism, (c) for the relationship between SDO and Choice (Footbridge dilemma) as mediated by altruism. SDO = social dominance orientation; TL = Trolley; FB = Footbridge. * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

The present study aimed to examine the association between authoritarian attitudes and moral decision-making in moral dilemmas pitting utilitarian gains against aversive harmful actions, and the potential mediating role of dispositional altruism. Following the procedure advocated by Tassy et al. (2013), our experimental design dissociated two moral scenarios (impersonal dilemma and personal dilemma) as well as two aspects of moral evaluation (“evaluative judgment” and “choice of hypothetical action”). Participants were more likely to endorse utilitarian outcomes in the impersonal dilemma (as compared to the personal dilemma), suggesting that the direct nature of the action may have influenced participants’ moral judgments. More importantly, our results highlight that high SDO scorers (i) judged both impersonal and personal harms as less permissible, (ii) favored the utilitarian options in both personal and impersonal dilemmas, (iii) reported less guilt in case of utilitarian actions. Thus, high SDO scorers were more likely to endorse utilitarian outcomes regardless of whether the dilemmas entailed direct or indirect harms, and less likely to experience guilt if the proposed action is performed. Finally, altruism fully mediated the association (i) between SDO and evaluative judgment in the impersonal dilemma (Trolley dilemma), (ii) between SDO and both evaluative judgment and choice of hypothetical action in the personal dilemma (Footbridge dilemma).

Authoritarianism and utilitarianism

The most likely culprit in terms of a common thread linking utilitarian decision making and SDO is a low aversion to harming others. This is in line with previous research on the antecedents of SDO showing that this attitude is related to tough mindedness (e.g., Duckitt, 2001), antisocial personality traits such as psychopathy (e.g., Hodson et al., 2009), low concern about fairness and harm avoidance (e.g., Milojev et al., 2014), low concern for

the others' well-being (e.g., Sidanius & Pratto, 1999) and less emphatic neural response (Chiao et al., 2009). Further support for the idea that increased utilitarianism seen among high SDO individuals is due to a deficit in emotional processing comes from the wealth of research showing a significant relationship between emotional experience and moral functioning. A large range of studies provides evidence (i) that harm aversion is a fundamental aspect of moral functioning (e.g., Crockett et al., 2010), (ii) that aversive emotional reactions to harmful acts rely on the ability to perceive, share and understand the others' affective states (e.g., Decety & Lamm, 2009). Behavioral and brain imaging studies support the notion that care-based moral reasoning and prosocial behavior rely on sympathetic responses to the others' pain and distress (e.g., Eisenberg & Eggum, 2009). Lesions studies have shown that damage to the "affective moral brain" results in a significant deficit in moral judgment (e.g., Koenigs et al., 2007; Funk & Gazzaniga, 2009; Moll & de Oliveira-Souza, 2007).

As mentioned above, participants were less likely to endorse utilitarian outcomes in the personal (vs. impersonal) dilemma. This is in line with prior work demonstrating that increased affective proximity leads to increased emotional responding to the others' distress (e.g., Bourgeois & Hess, 2008; Mermillod & Lepage, 2016) and decreased utilitarianism (e.g., Thomas, Croft, & Tranel, 2011). However, our results suggest that high SDO scorers have not been influenced by the affective proximity with the potential victim. This finding is consistent with research showing (i) that individuals low in emotional responsiveness (such as those high in psychopathy) are biased toward utilitarianism, (ii) that emotional responding cannot be potentiated in individuals high in psychopathic traits (e.g., Tassy et al., 2013).

Dispositional altruism and guilt

The mediating role of dispositional altruism provides additional support for the idea that the association between SDO and utilitarianism is driven by low concern for the others.

Recent research provides strong evidence that altruism is primarily driven by other-oriented feelings of concern (vs. self-oriented feelings of personal distress; e.g., FeldmanHall, Dalgleish, Evans, & Mobbs, 2015), and Kahane et al. (2015) first found a negative association between utilitarian judgments and altruist concern for distant strangers.

Our findings also highlight that ideology plays a role in the proclivities to experience guilt. This high-order emotion has often been operationalized as a response involving concern about others' feelings (e.g. Tangney, 1998) and is believed to motivate moral behavior (e.g., Hoffman, 1998). The guilty individual "accepts responsibility for a behavior that violates internalized standards or causes another's distress and desires to make amends or punish the self" (Eisenberg, 2000). Our results align with studies suggesting that the anticipation of the social-emotional consequences of an action (e.g., guilt) is critical for reasoning about moral dilemmas (e.g., Koenigs et al., 2007). Lesions studies have shown that the ventromedial prefrontal cortex (VMPFC) damage is associated with a lack of concern for others, a reduction in the display of high-order emotions such as guilt, and increased utilitarianism (Koenigs et al., 2007). According to Thomas et al. (2011), given the VMPFC's role in anticipating the consequences of an action and encoding the socio-emotional value of those consequences, the anticipation of serious social-emotional consequences of the proposed action is required to reason normally about the Trolley-type dilemmas. This anticipation triggers emotional responses such as guilt and "abnormal judgment pattern results when this emotional information is not properly activated and factored into the decision" (Thomas et al., 2011). The idea that anticipation triggers emotions that influence moral decision-making is consistent with research showing that anticipating an emotional experience leads to a partial simulation of that experience (e.g., Niedenthal, 2007).

All in all, our results provide evidence for substantial differences in moral reasoning based on ideology and for a harm aversion account of moral judgment. Such findings could have societal implications. Members of dominant groups are known to exhibit higher levels of SDO and situations do not affect the relative dominance of these individuals (Levin, 1996; Pratto et al., 1994; Sidanius, Levin, Liu, & Pratto, 2000). Thus, the extent to which a person is able to share the others' pain or misfortune may contribute to the policy formulation and processes, and the fact that members of dominant groups could be more likely to favor decisions resulting in harm to others could have detrimental social consequences (e.g., Stuckler & Basu, 2013).

Conclusion

The present study provides evidence for a relation between ideology and responses to moral dilemmas pitting utilitarian considerations against aversive acts of causing harm to others. High SDO scores predicted greater proportion of utilitarian responses for both judgment and choice, and lower anticipation of the social-emotional consequences of the proposed action (i.e., guilt), regardless of whether the dilemma was impersonal or personal. Lower altruistic concern fully mediated the association between SDO and utilitarian decision. This is consistent with research suggesting that many of our highest moral decisions are based on spontaneous, emotional intuitions and “gut-feelings”. All in all, our results provide a coherent harm aversion account of the relationship between ideology and moral judgment.

References

- Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg: University of Manitoba Press.
- Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.
- Altemeyer, B. (1998). The other “authoritarian personality.” In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 47-92). San Diego: Academic Press.
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11*(2), 150-166. doi:10.1177/1088868306294907
- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*(1), 154-161. doi:10.1016/j.cognition.2011.05.010
- Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.
- Baumert, A., Schlösser, T., & Schmitt, M. (2014). Economic games: A performance-based assessment of fairness and altruism. *European Journal of Psychological Assessment, 30*(3), 178-192. doi:10.1027/1015-5759/a000183
- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences, 11*(9), 387-392. doi:10.1016/j.tics.2007.07.003
- Blair, R. J. R., Mitchell, D., & Blair, K. (2005). *The Psychopath: Emotion and the Brain*. Oxford, UK: Blackwell.
- Bobbio, N. (1996). *Left & right: The significance of a political distinction*. Chicago: University of Chicago Press.

- Boies, K., Yoo, T.-Y., Ebacher, A., Lee, K., & Ashton, M. C. (2004). Psychometric properties of the French and Korean versions of the HEXACO Personality Inventory. *Educational and Psychological Measurement, 64*, 992-1006. doi:10.1177/0013164404267277
- Bostyn, D. H., Roets, A., & Van Hiel, A. (2016). Right-wing attitudes and moral cognition: Are right-wing authoritarianism and social dominance orientation related to utilitarian judgment? *Personality and Individual Differences, 96*, 164-171. doi:10.1016/j.paid.2016.03.006
- Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological Psychology, 77*(3), 343-352. doi:10.1016/j.biopsycho.2007.11.008
- Carney, D., & Mason, M. F. (2010). Decision making and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology, 46*(4), 668-671. doi:10.1016/j.jesp.2010.02.003
- Chiao, J. Y., Mathur, V. A., Harada, T., & Lipke, T. (2009). Neural basis of preference for human social hierarchy versus egalitarianism. *Annals of the New York Academy of Sciences, 1167*, 174-81. doi:10.1111/j.1749-6632.2009.04508.x
- Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences, 17*(8), 363-366. doi:10.1016/j.tics.2013.06.005
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences, 107*(40), 17433-17438. doi:10.1073/pnas.1009396107
- de Waal, F. B. M. (2003). Morality and the Social Instincts: Continuity with the Other Primates. In G. B. Peterson (Ed.), *The Tanner Lectures on Human Values* (pp. 1-40).

- Salt Lake City, UT: University of Utah Press. Retrieved from http://tannerlectures.utah.edu/_documents/a-to-z/d/deWaal_2005.pdf
- Decety, J., & Lamm, C. (2009). Empathy versus Personal Distress: Recent Evidence from Social Neuroscience. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 199-213). Cambridge, MA: MIT Press.
- Duarte, S., Dambrun, M., & Guimond, S. (2004). La dominance sociale et les “mythes légitimateurs”: validation d’une version française de l’échelle d’orientation à la dominance sociale. *Revue Internationale de Psychologie Sociale*, *17*(4), 97-126.
- Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 33, pp. 41-113). New York: Academic Press.
- Duckitt, J., & Sibley, C. G. (2010). Personality, ideology, prejudice, and politics: A dual process motivational model. *Journal of Personality*, *78*(6), 1861-1894. doi:10.1111/j.1467-6494.2010.00672.x
- Duke, A., & Bègue, L. (2015). The Drunk Utilitarian: Blood Alcohol Concentration Predicts Utilitarian Responses in Moral Dilemmas. *Cognition*, *134*, 121-127. doi:10.1016/j.cognition.2014.09.006
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, *51*(1), 665-697. doi:10.1146/annurev.psych.51.1.665
- Eisenberg, N., & Eggum, N. D. (2009). Empathic Responding: Sympathy and Personal Distress. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71-83). Cambridge, MA: MIT Press.
- Eisenberg, N., & Morris, A. S. (2001). The origins and social significance of empathy-related responding. A review of empathy and moral development: implications for caring and

- justice by M. L. Hoffman. *Social Justice Research*, 14(1), 95-120.
doi:10.1023/A:1012579805721
- Federico, C. M., Weber, C. R., Ergun, D., & Hunt, C. (2013). Mapping the connections between politics and morality: The multiple sociopolitical orientations involved in moral intuition. *Political Psychology*, 34(4), 589-610. doi:10.1111/pops.12006
- FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. *Neuroimage*, 105, 347-356. doi:10.1016/j.neuroimage.2014.10.043
- Ferguson, T. J., & Stegge, H. (1998). Measuring guilt in children: a rose by any other name still has thorns. In J. Bybee (Ed.), *Guilt and Children* (pp. 19–74). New York: Academic Press.
- Funk, C. M., & Gazzaniga, M. S. (2009). The functional brain architecture of human morality. *Current Opinion in Neurobiology*, 19(6), 678-681.
doi:10.1016/j.conb.2009.09.011
- Graham, J., Haidt, J., & Nosek, B. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029-46.
doi:10.1037/a0015141
- Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The neuroscience of morality* (pp. 35–79). Cambridge, MA: MIT Press.
- Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, 45(3), 581-584. doi:10.1016/j.jesp.2009.01.003
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293(5537), 2105-2108. doi:10.1126/science.1062872

- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*(1), 98-116.
doi:0.1007/s11211-007-0034-z
- Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences, 37*, 297-350.
doi:10.1017/S0140525X13001192
- Hodson, G., Hogg, S. M., & MacInnis, C. C. (2009). The role of "dark personalities" (narcissism, Machiavellianism, psychopathy), Big Five personality factors, and ideology in explaining prejudice. *Journal of Research in Personality, 43*(4), 686-690.
doi:10.1016/j.jrp.2009.02.005
- Hoffman, M. L. (1998). Varieties of empathy-based guilt. In J. Bybee (Ed.), *Guilt and Children* (pp. 91-112). New York: Academic Press.
- Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. New York: Cambridge University Press.
- Høgh-Olesen, H. (2010). Homo sapiens - Homo socius: A comparative analysis of human mind and kind. In H. Høgh-Olesen (Ed.), *Human morality and sociality: Evolutionary and comparative perspectives* (pp. 235-271). Basingstoke: Palgrave Macmillan.
- Huxley, T. H. (1989 [1894]). *Evolution and Ethics*. Princeton: Princeton University Press.
- Jost, J. T., & Amodio, D. M. (2012). Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion, 36*(1), 55-64.
doi:10.1007/s11031-011-9260-7
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129*(3), 339-375.
doi:10.1037/0033-2909.129.3.339

- Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, *134*, 193-209. doi:10.1016/j.cognition.2014.10.005
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*(7138), 908-911. doi:10.1038/nature05631
- Kugler, M., Jost, J. T., & Noorbaloochi, S. (2014). Another look at Moral Foundations Theory: Do Authoritarianism and Social Dominance Orientation Explain Liberal-Conservative differences in "Moral" Intuitions? *Social Justice Research*, *27*(4), 413-431. doi:10.1007/s11211-014-0223-5
- Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO Personality Inventory: Two new facet scales and an observer report form. *Psychological Assessment*, *18*(2), 182-191. doi:10.1037/1040-3590.18.2.182
- Levin, S. L. (1996). *A social psychological approach to understanding intergroup attitudes in the United States and Israel*. (Unpublished doctoral dissertation). UCLA, California.
- Maner, J. K., Luce, C. L., Neuberg, S. L., Cialdini, R. B., Brown, S., & Sagarin, B. J. (2002). The effects of perspective taking on helping: Still no evidence for altruism. *Personality and Social Psychology Bulletin*, *28*(11), 1601-1610. doi:10.1177/014616702237586
- McFarland, S. G. (2005). On the eve of war: Authoritarianism, social dominance and American students' attitudes toward attacking Iraq. *Personality and Social Psychology Bulletin*, *31*(3), 360-367. doi:10.1177/0146167204271596
- Mermillod, M., & Lepage, J. (2016). Embodied Emotion: the Functional Mystery of Embodying Emotions. In Y. Coello & M. H. Fischer (Eds.), *Foundations of Embodied Cognition, Perceptual and Emotional Embodiment (Vol. 1)*. Hove: Psychology Press.

- Milojev, P., Osborne, D., Greaves, L. M., Bulbulia, J., Wilson, M. S., Davies, C. L., Liu, J. H., & Sibley, C. G. (2014). Right-wing authoritarianism and social dominance orientation predict different moral signatures. *Social Justice Research, 27*(2), 149-174. doi:10.1007/s11211-014-0213-7
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences, 11*(8), 319-321. doi:10.1016/j.tics.2007.06.001
- Montoya, E. R., Terburg, D., Bos, P. A., Will, G. J., Buskens, V., Raub, W., & Van Honk, J. (2013). Testosterone administration modulates moral judgments depending on second-to-fourth digit ratio. *Psychoneuroendocrinology, 38*(8), 1362-1369. doi:10.1016/j.psyneuen.2012.12.001
- Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*, 1002-1005. doi:10.1126/science.1136930
- Passini, S. (2008). Exploring the multidimensional facets of authoritarianism: Authoritarian aggression and social dominance orientation. *Swiss Journal of Psychology, 67*(1), 51-60. doi:10.1024/1421-0185.67.1.51
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science, 36*(1), 163-177. doi:10.1111/j.1551-6709.2011.01210.x
- Penner, L. A., & Finkelstein, M. A. (1998). Dispositional and structural determinants of volunteerism. *Journal of Personality and Social Psychology, 74*(2), 525-537. doi:10.1037/0022-3514.74.2.525
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*(4), 741-763. doi:10.1037/0022-3514.67.4.741

- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717-731. doi:10.3758/BF03206553
- Schwartz, S. H., Caprara, G. V., & Vecchione, M. (2010). Basic personal values, core political values, and voting: A longitudinal analysis. *Political Psychology*, *31*(3), 421-452. doi:10.1111/j.1467-9221.2010.00764.x
- Sidanius, J., Levin, S., Liu, J., & Pratto, F. (2000). Social dominance orientation, anti-egalitarianism, and the political psychology of gender: An extension and cross-cultural replication. *European Journal of Social Psychology*, *30*(1), 41-67. doi:10.1002/(SICI)1099-0992(200001/02)30:1<41::AID-EJSP976>3.0.CO;2-O
- Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.
- Singer, P. (2005). Ethics and intuitions. *Journal of Ethics*, *9*(3-4), 331-352. doi:0.1007/s10892-005-3508-y
- Staub, E. (1974). Helping a distressed person: Social, personality, and stimulus determinants. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 293-341). New York: Academic Press.
- Stuckler, D., & Basu, S. (2013). *The Body Economic: Why Austerity Kills*. London: Basic Books.
- Tangney, J. P. (1998). How does guilt differ from shame? In J. Bybee (Ed.), *Guilt and Children* (pp. 1-17). New York: Academic Press.
- Tassy, S., Deruelle, C., Mancini, J., Leistedt, S., & Wicker, B. (2013) High levels of psychopathic traits alters moral choice but not moral judgment. *Frontiers in Human Neuroscience*, *7*, 229. doi:10.3389/fnhum.2013.00229

- Thomas, B. C., Croft, K. E., & Tranel, D. (2011). Harming kin to save strangers: further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, 23(9), 2186-2196. doi:10.1162/jocn.2010.21591
- Tost, H., & Meyer-Lindenberg, A. (2010). I fear for you: A role in serotonin in moral behaviour. *Proceedings of the National Academy of Sciences*, 107(40), 17071-17072. doi:10.1073/pnas.10125451
- Zettler, I., & Hilbig, B. E. (2010). Attitudes of the selfless: Explaining political orientation with altruism. *Personality and Individual Differences*, 48(3), 338-342. doi:10.1016/j.paid.2009.11.002

Idéologie, altruisme et utilitarisme moral

L'étude 1 a permis de mettre en évidence une relation entre SDO et préférence pour les options utilitaristes dans les deux versions du « dilemme du tramway ». Cette relation est médiatisée par un moindre altruisme dispositionnel. La SDO est également négativement associée au sentiment de culpabilité en cas de réalisation de l'action proposée. Dans leur ensemble, ces résultats suggèrent que cette dimension de l'autoritarisme est associée à une moindre aversion pour l'atteinte à autrui dans une situation de dilemme moral. Par contraste, l'absence de relation entre RWA et utilitarisme moral apporte un premier soutien à l'idée voulant que l'influence du RWA sur l'OD ne soit pas la conséquence d'un déficit de résonance empathique pour moduler une évaluation utilitariste du dilemme social induit par le paradigme de Milgram. L'objectif de l'étude 2 était d'explorer la relation entre autoritarisme et tonus vagal.

3.2. Autoritarisme et Tonus vagal

Objectifs et présentation de l'étude

La recherche en psychologie politique s'est essentiellement inscrite dans un paradigme voulant que les différences interindividuelles soient socialement déterminées. Dans ce contexte, peu d'auteurs se sont intéressés aux aspects biologiques comme facteur prédictif et explicatif de l'autoritarisme. Or, le comportement humain est déterminé la fois par des facteurs environnementaux et biologiques, et ce même dans ses aspects politiques (e.g., Hatemi & McDermott, 2011). La littérature suggérant que le RWA soit l'expression dans la sphère sociale d'un cluster « rigidité cognitive, anxiété et attitude défensive » (e.g., Jugert & Duckitt, 2009), nous avons fait l'hypothèse qu'un moindre tonus vagal serait associé à des scores supérieurs à l'échelle RWA. Afin de tester cette hypothèse, la HRVb des participants ($n = 119$) a été enregistrée sur une période de 5 minutes (temps standard). Ces derniers ont ensuite répondu à un questionnaire incluant les versions françaises des échelles RWA et SDO.

Manuscrit 2⁷

⁷ Manuscrit soumis pour publication auprès de la revue *Motivation and Emotion*.

Cardiac Vagal Tone and Political Attitudes

Johan Lepage¹, Laurent Bègue¹, Michael Dambrun², & Martial Mermillod^{3,4}

¹ Université Grenoble Alpes, Laboratoire InterUniversitaire de Psychologie

² Université Blaise Pascal, Laboratoire de Psychologie Sociale et Cognitive

³ Université Grenoble Alpes, Laboratoire de Psychologie et NeuroCognition

⁴ Institut Universitaire de France (Paris, France)

Corresponding author

Johan Lepage, PhD Candidate: Laboratoire InterUniversitaire de Psychologie, UFR SHS,
1251 avenue Centrale, BP 47, 38040 Grenoble Cedex 9, France; E-mail:
johan.lepage@gmail.com, Phone: (+33 4) 76 82 58 92, Fax: (+33 4) 76 82 56 65

Abstract

Right-wing authoritarianism (RWA) and social dominance orientation (SDO) both predict a large range of sociopolitical phenomena such as prejudice, patriotism, militarism, or support for capital punishment. A great deal of research has been devoted to explaining the origins and manifestations of political attitudes but only few studies considered biological differences as an important explanatory and predictive factor. In the present study, we investigated the association between cardiac vagal tone (VT) and political attitudes. A host of studies causally relate decreased VT at rest with increased threat perceptions, stress and decreased prosocial tendencies. We assessed the strength of VT (as indexed by resting heart rate variability) of 119 healthy adults before they completed a survey assessing RWA and SDO. Results show that reduced VT at rest predicts higher RWA scores and SDO scores. Implications of such results for research on political attitudes are discussed in relation to the polyvagal theory and the model of neurovisceral integration.

Keyword

Right-wing authoritarianism, social dominance orientation, cardiac vagal tone, threat, stress vulnerability

Introduction

Political psychology research suggests that right-wing authoritarianism (RWA) and social dominance orientation (SDO) are the core dimensions of ideology (e.g., Duckitt, 2001). These attitudes both predict a wide range of sociopolitical phenomena such as generalized prejudice, intolerance, right-wing political party preference, nationalism, patriotism, militarism, support for capital punishment, and generally punitive attitudes (e.g., Altemeyer, 1998; Duckitt, 2001; Kteily, Sidanius, & Levin, 2011). Across human cultures, SDO and RWA appear to be two relatively stable individual-difference orientations over the lifecourse (e.g. Hatemi et al., 2009; Ludeke & Krueger, 2013). A great deal of research has been devoted to explaining the origins and manifestations of ideology, and traditional explanations rest principally on environmental stimuli as ultimate cause of political attitudes and behaviors (e.g., Campbell, Converse, Miller, & Stokes, 1960; Wendt, 1999). As emphasized by Hatemi and McDermott (2011), current models “tell only a portion of the human story” by excluding human biological needs and drives. The present study aimed to advance our understanding of the ways in which physiology may influence the development and manifestations of political attitudes. The polyvagal theory (e.g., Porges, 1995, 2007) and the model of neurovisceral integration (e.g., Thayer & Lane, 2000; Thayer, Hansen, Saus-Rose, & Johnsen, 2009) both causally relate cardiac vagal tone (VT) with self-regulation processes and prosocial tendencies. This work aimed to explore the association between political attitudes and VT.

Right-wing authoritarianism

First conceptions claimed (i) that authoritarianism is the direct expression of a personality dimension, (ii) that negative emotions and cognitive rigidity are the core psychological factors that enhance authoritarian beliefs (Adorno, Frenkel-Brunswik, Levinson & Sanford, 1950; Allport, 1954; Rokeach, 1954; Wilson, 1973). As its predecessors,

Altemeyer (1981, 1988, 1998) conceptualized authoritarianism as a personality dimension. The items of his RWA scale capture the covariation of three dimensions: (i) authoritarian submission (a submissive, uncritical attitudes toward authorities), (ii) authoritarian aggression (tendency to condemn, reject, and punish people who violate conventional values), and (iii) conventionalism (rigid adherence to conventional middle-class values). A large range of studies support the validity of the RWA scale as a measure of these three “traits” (e.g., Duckitt, 2013). Experimental and longitudinal research on the antecedents of RWA demonstrates the influence of social threat (e.g., political and economic crises, insecurity): (i) a perceived threat increases RWA at both individual and societal levels (e.g., Doty, Peterson, & Winter, 1991; McCann, 1997; Rickert, 1998; Rohan & Zanna, 1996; Sales, 1972; Sales & Friend, 1973), (ii) high RWA individuals are more reactive to threat (Feldman & Stenner, 1997; Greenberg et al., 1990; Lavine, Sweeney, & Wagne, 1999; Lavine, Lodge, & Freitas, 2005), (iii) experimental manipulation of threat influence RWA scores (Duckitt & Fisher, 2003; Jugert & Hiemisch, 2005). Research also suggests a relation between several personality traits and RWA. For instance, studies found that RWA is positively related to conscientiousness and negatively related to openness (for a review see Sibley & Duckitt, 2008). Others studies demonstrate that RWA is associated with need for order, closure and structure (e.g., Cornelis & Van Hiel, 2006; Kossowska & Van Hiel, 2003; Van Hiel, Pandelaere, & Duriez, 2004).

Social dominance orientation

SDO has been defined as a “general attitudinal orientation toward intergroup relations, reflecting whether one generally prefers such relations to be equal, versus hierarchical” (Pratto, Sidanius, Stallworth & Malle, 1994). Research shown that SDO predicts a large range of authoritarian phenomena similar to those predicted by RWA (e.g., Pratto, Sidanius, &

Levin, 2006; Sidanius & Pratto, 1999). Experimental research on the antecedents of SDO has demonstrated that high-status identity primes and changes in relative group status can increase individuals SDO scores (e.g., Guimond, Dambrun, Michinov, & Duarte, 2003; Huang & Liu, 2005; Schmitt, Branscombe, & Kappen, 2003). For example, Schmitt et al. (2003) found that students exhibited higher SDO scores after being told that their university had more (*vs.* fewer) resources than other universities. Research also demonstrates that SDO is positively correlated with several antisocial personality traits such as Psychopathy and Machiavellianism (e.g., Hodson, Hogg, & MacInnis, 2009) and negatively related to concern for the others' well-being, fairness and harm avoidance (Federico, Weber, Ergun, & Hunt, 2013; Milojev et al., 2014; Pratto et al., 1994; Sidanius & Pratto, 1999; see also Kugler, Jost & Noorbaloochi, 2014). Others studies have shown that SDO bolstered support for war by reducing concern for its human costs (e.g., McFarland, 2005).

Motivational approach of authoritarianism

According to the Dual Process Motivational (DPM) model (e.g., Duckitt & Sibley, 2010), RWA and SDO are two distinct dimensions of ideological attitudes with each expressing motivational goals and values made chronically salient for individuals by their worldviews and personality traits (e.g., Duckitt, 2001). High RWA expresses the motivational goal and value of establishing or maintaining collective security, which are made chronically salient by the belief that the social world is inherently dangerous and threatening. High SDO expresses the motivational goals and values of power, dominance, and superiority over others, chronically salient because of a view of the world as a ruthlessly competitive jungle in which the strong win and the weak lose. For RWA, the predisposing personality dimension is social conformity (i.e., low Openness to experience and high Conscientiousness), while SDO stems

from the underlying personality dimension of tough *vs.* tender mindedness (i.e., low Agreeableness).

Unfortunately, biological independent variables are extremely rare in research on RWA and SDO, and biological theories rarely used to generate hypotheses about political attitudes. A widely-held assumption is that political attitudes and behaviors are entirely the product of environmental forces (e.g., parental socialization, idiosyncratic experiences). But as emphasized by Hatemi and McDermott (2011), “who we are and what we do is the product of environmental *and* biological factors” and human political behavior “does not transcend biology even as it is not determined by biology”. Because a large amount of research using subjective self-report suggests that RWA is strongly related to threat sensitivity, anxiety, cognitive rigidity and defensiveness (e.g., Duckitt & Sibley, 2010), VT seems to qualify as a potential predictor of RWA. A plethora of studies demonstrates that VT reflects self-regulation processes and represents a robust biomarker of threat perceptions and stress.

Cardiac vagal tone as an index of self-regulation processes

The autonomic nervous system (ANS) is subdivided into an excitatory sympathetic nervous system (SNS) and an inhibitory parasympathetic nervous system (PNS). When an individual is threatened, the activity of the SNS becomes dominant, triggering a suite of physiological responses to aid in adapting to the challenge (e.g., increased heart rate). In a safe context, the PNS is dominant and maintains a lower degree of physiological arousal (e.g., decreased heart rate). A flexible ANS allows for rapid modulation of physiological states in accordance with situational demands. In contrast, autonomic rigidity results in a lessened capacity to alter physiological responses in synchrony with environmental changes (e.g., Appelhans & Luecken, 2006). The polyvagal theory (e.g., Porges, 1995, 2007) causally relates autonomic flexibility with emotion-related self-regulation and prosocial tendencies.

Polyvagal theory

The polyvagal theory (e.g., Porges, 1995, 2007) articulates three phylogenetically-ordered neural circuits regulating the heart: (i) the “social engagement system” is dependent upon the functions of the myelinated vagus “which serves to foster calm behavioral states by inhibiting the sympathetic influences to the heart” (Porges, 2007); (ii) the “fight-or-flight system” is dependent on the functioning of the SNS; (iii) the “immobilization system” (the most phylogenetically primitive component) is dependent on the unmyelinated vagus. In safe contexts, there is an increased influence of the myelinated vagus on the cardiac pacemaker that slows the heart and inhibits the fight/flight mechanisms of the SNS. The two more primitive neural circuits serve to ensure survival in life-threatening contexts through defensive strategies (i.e., fight–flight and death-feigning behaviors). Under resting conditions, the myelinated vagus acts as a “brake”, firing a rapid and continuous signal that slows the heart.

The polyvagal theory also states that a neural process termed “neuroception” continuously evaluates risk and modulates vagal activity via higher brain structures. An invalid neuroception of danger is believed to contribute to maladaptive physiological reactivity and the expression of the defensive behaviors associated with negative emotional states (e.g., fear, anxiety; Porges, 2007).

The strength of the tonic vagal influence is commonly assessed by examining heart rate variability (HRV) at rest. Following the polyvagal perspective, higher resting HRV should be related to higher self-regulatory control and prosocial behaviors. Consistently, studies found that higher resting HRV predicted greater regulatory control, decreased negative emotional arousal in response to stressors, sympathy, compassion and helping behavior (e.g., Fabes & Eisenberg, 1997; Stellar, Cohen, Oveis, & Keltner, 2015). By contrast, lower resting

HRV has been related to personal distress, psychosocial stress, social anxiety and defensiveness (e.g., Movius & Allen, 2005; Porges, 1992; Thayer & Lane, 2007). More generally, a “ host of findings suggest that human subjects with greater vagal influence may be dispositionally more open to experience and better able to self-regulate” (Movius & Allen, 2005).

Model of neurovisceral integration

The model of neurovisceral integration (e.g., Thayer & Lane, 2000; Thayer et al., 2009) states that the neural networks implicated in cognitive and emotion self-regulation (e.g., ventromedial prefrontal cortices, anterior cingulate, insula, amygdala) are also involved in the control of cardiac activity. In safe contexts, the prefrontal cortex is believed to exert an inhibitory control over sympatho-excitatory subcortical circuits (e.g., Thayer et al., 2009). In threatening contexts, the prefrontal inhibitory regulation diminishes and sympatho-excitatory subcortical circuits produce default threat responses (e.g., Park, Vasey, Van Bavel, & Thayer, 2013). Consistently, brain imaging studies found that reduced prefrontal activity is responsible for hyperactive subcortical activity, prolonged activation of defensive behaviors and hyper-vigilance (e.g., Thayer et al., 2009; Park et al., 2013). Research also provides evidence that the functioning of prefrontal-subcortical inhibitory circuits is linked with the heart via the vagus nerve. For instance, studies found that lower VT at rest is associated with hypoactive prefrontal regulation, hyperactive subcortical structures, maladaptive cognitive and emotion regulation, increased negativity bias (i.e., the tendency to prioritize negative information over positive) and decreased willingness to approach positive novel objects (see Thayer, Ahs, Fredrikson, Sollers, & Wager, 2012). In this perspective, VT reflects inhibitory prefrontal processes and represents a useful index of threat perceptions and therefore of stress.

As mentioned above, previous political psychology research suggests that RWA is related to anxiety, cognitive rigidity, threat sensitivity and defensiveness. Thus, we hypothesized that lower VT at rest would predict higher RWA scores.

Material and method

Participants

One hundred nineteen healthy adults (83 females, $M_{age} = 22.15$; $SD = 2.76$) volunteered for the experiment in exchange for course credit or 10€. Participants were recruited through newspaper advertisements, public notices, flyers and direct email solicitation. Participants were French or bilingual in French. They provided written informed consent before the participation. The study was part of a global project reviewed and approved by the University human ethics committee from Grenoble, France (Grenoble ethics committee notice number 2013-07-04-23).

To be eligible, participants had to be aged between 18 and 50 years, with a normal or normal-to-corrected vision, explicitly reported an absence of psychiatric, neurologic, hormonal, or cardio-vascular disease, and with no medical treatment (with the exception of contraception). Smoking, energizing drinks (e.g., coffee, tea, etc.) and psychotropic substances (e.g. alcohol, cannabis, etc.) were prohibited to each participant the day of the experiment. They had also to avoid eating or drinking (water was allowed) the 2 hours preceding the experiment in order to limit the influence of digestion on autonomic functioning but they had to eat in the morning (more than 2 hours before the experiment) in order to avoid fasting states (Short term HRV measurement can be biased by the digestion of food since viscera are innervated by the autonomic nervous system; Heathers, 2014; Iorfino, Alvares, Guastella, & Quintana, 2015; Quintana & Heathers, 2014). Finally, they had to avoid water

consumption 30 minutes before the experiment and were asked to empty their bladder before starting the experiment.

R-to-R measurement

HRV measurements include time domain and frequency domain methods. In the time-domain, the most popular estimate of HRV is RMSSD (root mean square of successive R-R differences), which is calculated as the standard deviation of the beat-to-beat intervals. A popular frequency-domain technique to estimate HRV involves decomposing heart period variance into different frequency bands using Fourier transformations. The high frequencies (HF-HRV, 0.15-0.4 Hz) are thought to represent primarily vagal influence, and as such parasympathetic activity (e.g., Montano, Porta, & Malliani, 2001). The low frequencies (LF-HRV, 0.04-0.15 Hz) are believed to represent the influence of both sympathetic and parasympathetic activities (see Negrao, Bipath, van der Westhuizen, & Viljoen, 2011 for a summary). Several studies demonstrate that RMSSD and HF-HRV are positively correlated (e.g., Otzenberger et al., 1998).

Resting HRV was recorded during 10 minutes. Three Ag-AgCl electrodes were attached as follows: one electrode (Negative) on right clavicle and one electrode (Positive) on the left 4th or 5th rib mid-clavicular to axillary line, with the Ground on right 4th or 5th rib mid-clavicular to axillary line. ECG (sampling rate 1000 Hz) was recorded using Single BIO Amp hardware in conjunction with PowerLab 16/32 hardware and LabChart 8 software (ADInstruments Ltd, UK).

Procedure

The experiment took place in a quiet and dimmed room; participants were installed in a comfortable seat. After being fully informed about the experiment, participants gave their

written consent for the participation. They were told that their cardiac activity will be recorded during 10 minutes and that they should try to relax. The 5 first minutes allowed participant to shift in a quiet state and HRV was recorded during the 5 following minutes. When this step ended, participants completed an online survey (Inquisit Web). RWA was assessed using 10 items ($\alpha = .77$) from the French translation of Altemeyer's 20-item (1988) RWA Scale (1 = *totally disagree* to 7 = *totally agree*). SDO was assessed using 10 items ($\alpha = .84$) from the French translation of the 16-items version (Duarte, Dambrun, & Guimond, 2004) of the SDO Scale (1 = *totally disagree* to 7 = *totally agree*).

Data analysis

Prior to analysis, signal was visually inspected and ectopic (abnormal) heart beats were removed using the LabChart Beat Classifier view. The HRV 2.0 Add-On for LabChart was used to calculate RMSSD (root mean square of successive R-R differences), HF-HRV_{n.u.} (0.15-0.4 Hz; normalized units) and LF-HRV_{n.u.} (0.04-0.15 Hz; normalized units). The HRV 2.0 Add-On module uses the Lomb Periodogram nonparametric method for spectral analysis which, unlike the Fast Fourier Transform method, can compute the spectral frequencies directly from the unevenly sampled tachogram. As the total power of the spectral signal is heterogeneous from person to person, it has been recommended that HF-HRV and LF-HRV data should be presented as normalized values (e.g., Pagani et al., 1986). The hypotheses were tested using simple linear regression (because of multicollinearity, the HRV parameters were not entered simultaneously). All our continuous predictors were centered before analysis.

Results

Descriptive statistics and Pearson's correlations are presented in Table 1. Results are presented in Table 2 and illustrated in Figure 1.

Table 1

Descriptive statistics and correlations between heart rate variability parameters and authoritarian attitudes

	1	2	3	4	5
1 RWA	—	.40**	-.41**	-.27**	.27**
2 SDO		—	-.15*	-.27**	.27**
3 RMSSD			—	.43**	-.44**
4 HF-HRV _{n.u.}				—	-.99***
5 LF-HRV _{n.u.}					—
<i>M</i> (<i>SD</i>)	2.98 (.86)	2.24 (.89)	43.48 (24.47)	42.16 (19.63)	57.85 (20.01)

Note. RWA = right-wing authoritarianism; SDO = social dominance orientation; RMSSD = root mean square of successive R-R differences; HF-HRV_{n.u.} = high frequencies (0.15-0.4 Hz) expressed in normalized units; LF-HRV_{n.u.} = low frequencies (0.04-0.15 Hz) expressed in normalized units. * $p < .05$ ** $p < .01$

Table 2

Linear regressions examining the associations between heart rate variability parameters and authoritarian attitudes

Criterion variables	Predictor variables	β	t	p	95% CI
RWA	RMSSD	-.41	-4.93	.000	[-.50, -.21]
	HF-HRV _{n.u.}	-.27	-3.03	.003	[-.38, -.08]
	LF-HRV _{n.u.}	.27	3.09	.002	[.08, .39]
SDO	RMSSD	-.15	-1.74	.08	[-.30, .02]
	HF-HRV _{n.u.}	-.27	-3.12	.002	[-.40, -.09]
	LF-HRV _{n.u.}	.27	3.08	.003	[.08, .40]

Note. RWA = right-wing authoritarianism; SDO = social dominance orientation; RMSSD = root mean square of successive R-R differences; HF-HRV_{n.u.} = high frequencies (0.15-0.4 Hz) expressed in normalized units; LF-HRV_{n.u.} = low frequencies (0.04-0.15 Hz) expressed in normalized units.

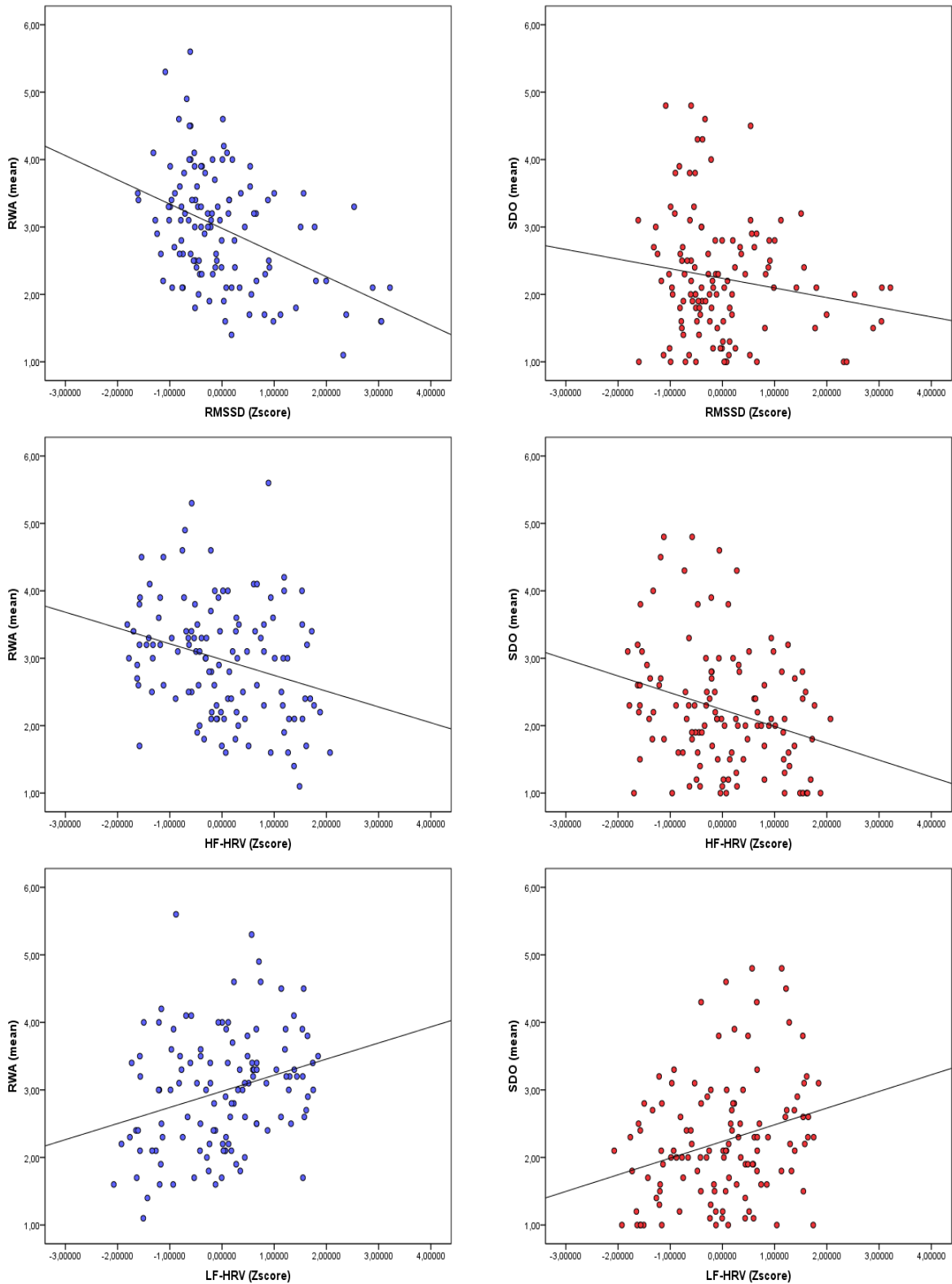


Figure 1. Scatter plots for the association between heart rate variability parameters and authoritarian attitudes. RWA = right-wing authoritarianism; SDO = social dominance orientation; RMSSD = root mean square of successive R-R differences; HF-HRV = high frequencies (0.15-0.4 Hz) expressed in normalized units; LF-HRV = low frequencies (0.04-0.15 Hz) expressed in normalized units.

Discussion

Our results revealed that lower VT at rest was associated with higher RWA scores and SDO scores. More precisely, we found that the markers of parasympathetic activity (RMSSD and HF-HRV) were associated with lower RWA scores and SDO scores. By contrast, higher LF-HRV (a marker of both sympathetic and parasympathetic activities) was associated with higher RWA scores and SDO scores. We didn't expect any relation between VT and SDO, but this finding is consistent with few recent studies examining the influence of realistic threat on SDO. Realistic threat is thought to involve a general concern for the ingroup's status and well-being (e.g., Stephan & Stephan, 2000). According to Morrison and Ybarra (2008) "one way in which group members may cognitively ready themselves to protect their ingroup's status and well-being is by exhibiting high SDO in response to threat". Consistently, the authors found that highly-identified group members (those who are particularly motivated to preserve their group's status and well-being) exhibited higher SDO scores in response to realistic threat (i.e., in response to perceived obstacles to the ingroup's status and general welfare). All in all, the present study suggests that low capacity to regulate attention and stress influence the experience and expression of political attitudes.

Seeing political attitudes through the lens of self-regulation has been spurred by research linking self-regulatory control and threat processing with ideology. Authoritarianism has been first theorized to involve threat sensitivity and cognitive rigidity. Allport (1954) argued that authoritarianism is characterized by insecurity and fearfulness. These feelings would cause to the authoritarian personalities (i) to a need structure, order and control in their social environments, and (ii) to react with punitive hostility to unconventionality, novelty and change. Rokeach (1954) preferred the concept of "dogmatism" defined as "a relatively closed cognitive organization of beliefs and disbeliefs about reality". Because of such a cognitive

configuration, dogmatic individuals should experience difficulty to deal with new information that challenges their existing beliefs. Dogmatic individuals should also feel hostility and reject persons and outgroups with dissimilar beliefs and values to their own. Wilson (1973) defined conservatism as a “generalized susceptibility to experiencing threat or anxiety in the face of uncertainty”. More recently, Jost and colleagues provided evidence that conservatism is associated with needs to manage uncertainty and threat (e.g., Jost, Glaser, Kruglanski, & Sulloway, 2003).

These theoretical propositions are comprehensive and thought-provoking but rely almost exclusively on subjective self-reports. With regard to implicit, subthreshold tendencies, RWA has been associated with automatic recognition of negative words suggesting threats to social order and group cohesion (Feldman, Lavine, Lodge & Verhulst, 2010). Other studies found that conservatives are more prone to perceive “emotionally ambiguous” faces as expressing a threatening or a dominant emotion (Vigil, 2010), and to perceive negative images more negatively (Hibbing, Smith, & Alford, 2014). Conservatives have also been found “to be quicker to focus on the negative, to spend longer looking at the negative, and to be more distracted by the negative” (Hibbing et al., 2014).

Research adopting a political neuroscience approach has begun to address the association between ideological outcomes and biological mechanisms. For instance, Oxley et al. (2008) found that conservatism was associated with larger skin conductance responses to threatening images (heightened skin conductance response reflects sympathetic activation), and stronger defensive startle reflex in response to unexpected auditory prompts. Using a “Go/No-Go task”, few electroencephalographic studies examined the association between ideology and conflict monitoring, an important aspect of cognitive self-regulation. Amodio, Jost, Master and Yee (2007) found that conservatism was associated with lower behavioral

accuracy on No-Go trials and lower conflict-related neural activity when response inhibition is required (see also Weissflog et al., 2010). Kanai et al. (2011) provide evidence that the functional correlate of ideology has a counterpart in brain structure. Using structural MRI, the authors found that conservatism was associated with lesser anterior cingulate cortex volume (a brain structure which play a critical role in cognitive flexibility) and larger right amygdala volume. Researchers have also turned to molecular genetics approaches, and Hatemi et al. (2011) found that genetic polymorphisms related to both cognitive flexibility and threat sensitivity predicted scores on a general attitudinal measure of liberalism-conservatism.

Hibbing et al. (2014) proposed that individual differences in negativity bias may underlie variations in ideology. Tritt, Peterson, and Inzlicht (2013) argued that emotional arousal, in general, and not negative valence, specifically, may underlie conservatism. In this perspective, the effect of negative emotion on the experience and expression of conservatism is due to engaging intensely arousing states. Amodio et al. (2007) stated that political orientation “reflects individual differences in the functioning of a general mechanism related to cognitive control and self-regulation”. Self-regulation has long been seen as part of political attitudes but almost no studies have used ANS functioning as a physiological index of self-regulation, thus leaving it unclear whether regulation is “impaired” in authoritarian attitudes per se. We believe that the study of political attitudes still needs a comprehensive and integrated conceptualization based on a general model of self-regulation. The use of current models of self-regulation such as the polyvagal theory and the model of neurovisceral integration may represent a significant step forward in understanding the causes and manifestations of political attitudes. The both approaches conceptualize VT as a useful index of threat perceptions and therefore of stress. Following a polyvagal perspective, our results suggest that autonomic rigidity (as indexed by low VT at rest) may be a core feature by which we may understand political attitudes. That is, an invalid neuroception of danger associated

with a reduced capacity to alter psychophysiological responses in synchrony with situational demands may underlie the psychological motivations to endorse authoritarian beliefs. The model of neurovisceral integration leads to similar conclusion, that is political attitudes may reflect individuals differences in a predisposition to chronic threat perception associated with dysregulated brain-peripheral integration (as indexed by low VT at rest). In other words, chronic difficulties in inhibiting the default response mode to threat, in down-regulating the sympatho-excitatory, cardio-acceleratory subcortical threat circuits, would cause to individuals to endorse authoritarian beliefs and punitive attitudes. These approaches also causally relate VT to social behaviors, thus providing insights on how political attitudes may affect concrete behavior.

Prejudice and discrimination

Why RWA and SDO causes prejudice and discrimination remain an open question (e.g., Kteily et al., 2011). Pratto et al. (1994) proposed that decreased inclusiveness and empathic concern among high SDO individuals play important roles. As mentioned previously, studies following a polyvagal perspective found that increased VT at rest is associated with calming and self-soothing (e.g., Porges, 2007), sympathy and compassion for the others in need (e.g., Stellar et al., 2015). Keltner, Kogan, Piff and Saturn (2014) states that “the relationship between vagus nerve activation and prosocial behavior is driven, in part, by a reduction in arousal, which enables a shift in attention to the person in need”. This is in line with research showing that low VT at rest is associated with rigid attentional bias toward threat-related stimuli (e.g., Park & Thayer, 2014). Thus, a reduced capacity to inhibit the sympatho-excitatory circuits responsible for default threat responses (i) may limit the ability for authoritarian individuals to challenge invalid threat perception, to pay attention to distress cues from the others and to display sympathetic responses, (ii) may promote punitive attitudes

and aggressive behaviors toward specific outgroups in ways favoring the perpetuation of conventional values and hierarchical arrangements amongst these groups. By contrast, the ability to inhibit the default response mode to threat and to focus on a broader set of social cues may increase flexibility by decreasing stress levels and facilitate prosocial tendencies (e.g., empathic concern, inclusiveness, tolerance).

Conclusion

To our knowledge, the present study is the first to explore the relationship between political attitudes and VT, suggesting that one organizing element of the many psychological and physiological traits related to political attitudes may be hyper-vigilance to threat-related stimuli and stress. The present study also highlights that biology “can help to account both for central behavioral tendencies and also for the astounding variation around these central tendencies” (Hatemi & McDermott, 2011). Because we did not experimentally provoke VT level, it will be essential to test causal associations between VT and political attitudes. Several techniques exist for experimentally inducing changes in VT, as is the case with pharmacological blockade and biofeedback training (e.g., Lehrer & Gevirtz, 2014). Interestingly, it has been proposed that higher VT could be responsible for improved social functioning and that increased social functioning could enhance VT (e.g., Kok & Fredrickson, 2010). These authors found an “upward spiral” in which VT “facilitates capitalizing on social and emotional opportunities and the resulting opportunistic gains, in turn, lead to higher vagal tone”. Thus, political attitudes and VT could reciprocally and prospectively predict one another.

References

- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Perseus Books.
- Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Canada: University of Manitoba Press.
- Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.
- Altemeyer, B. (1998). The other “authoritarian personality.” In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 47–92). Orlando, FL: Academic Press.
- Amodio, D. M., Jost, J. T., Master, S. L., & Yee, C. M. (2007). Neurocognitive correlates of liberalism and conservatism. *Nature Neuroscience*, *10*(10), 1246-1247.
doi:10.1038/nn1979
- Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, *10*(3), 229-240.
doi:10.1037/1089-2680.10.3.229
- Campbell, A. Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American Voter*. New York: Wiley.
- Cornelis, I., & Van Hiel, A. (2006). The impact of cognitive styles on authoritarianism based conservatism and racism. *Basic and Applied Social Psychology*, *28*(1), 37-50.
doi:10.1207/s15324834basp2801_4
- Duarte, S., Dambrun, M., & Guimond, S. (2004). La dominance sociale et les “mythes légitimateurs”: validation d’une version française de l’échelle d’orientation à la dominance sociale. *Revue Internationale de Psychologie Sociale*, *17*(4), 97-126.

- Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 33, pp. 41-113). New York: Academic Press.
- Duckitt, J. (2013). Introduction to the special section on authoritarianism in societal context: The role of threat. *International Journal of Psychology*, 48(1), 1-5. doi:10.1080/00207594.2012.738298
- Duckitt, J., & Fisher, K. (2003). The impact of social threat on worldview and ideological attitudes. *Political Psychology*, 24(1), 199-222. doi:10.1111/0162-895X.00322
- Duckitt, J., & Sibley, C. G. (2010). Personality, ideology, prejudice, and politics: A dual process motivational model. *Journal of Personality*, 78(6), 1861-1894. doi:10.1111/j.1467-6494.2010.00672.x
- Doty, R. G., Winter, D. G., Peterson, B. E., & Kimmelmeier, M. (1997). Authoritarianism and American students' attitudes about the Gulf War, 1990-1996. *Personality and Social Psychology Bulletin*, 23(11), 1133-1143.
- Fabes, R. A., & Eisenberg, N. (1997). Regulatory control and adults' stress-related responses to daily life events. *Journal of Personality and Social Psychology*, 73(5), 1107-1117. Retrieved from http://www.researchgate.net/profile/Nancy_Eisenberg/publication/13864629_Regulatory_control_and_adults%27_stress_related_responses_to_daily_life_events/links/02e7e539be941aade6000000.pdf
- Federico, C. M., Weber, C. R., Ergun, D., & Hunt, C. (2013). Mapping the connections between politics and morality: The multiple sociopolitical orientations involved in moral intuition. *Political Psychology*, 34(4), 589-610. doi:10.1111/pops.12006
- Feldman, S., Lavine, H., Lodge, M., & Verhulst, B. (2010). *Seeing Negative: Authoritarianism and Automatic Vigilance for Threatening Stimuli*. Paper presented at

the Annual Meeting of the International Society of Political Psychology, San Francisco, CA.

Feldman, S., & Stenner, K. (1997). Perceived Threat and Authoritarianism. *Political Psychology, 18*(4), 741-770. doi:10.1111/0162-895X.00077

Guimond, S., Dambrun, M., Michinov, N., & Duarte, S. (2003). Does social dominance generate prejudice? Integrating individual and contextual determinants of intergroup cognitions. *Journal of Personality and Social Psychology, 84*(4), 697-721. doi:10.1037/0022-3514.84.4.697

Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A., Veeder, M., Kirkland, S. & Lyon, D. (1990). Evidence for terror management theory II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview, *Journal of Personality and Social Psychology, 58*(2), 308-318. doi:10.1037/0022-3514.58.2.308

Hatemi, P. K., & McDermott, R. (2011). Evolution as a Theory for Political Behavior. In P. K. Hatemi & R. McDermott (Eds.), *Man is by Nature a Political Animal: Evolution, Biology, and Politics*. (pp. 13-46). Chicago, IL: University of Chicago Press.

Hatemi, P. K., Funk, C. L., Medland, S. E., Maes, H. M., Silberg, J. L., Martin, N. G., & Eaves, L. J. (2009). Genetic and Environmental Transmission of Political Attitudes Over a Life Time. *The Journal of Politics, 71*(3), 1141-1156. doi:10.1017/S0022381609090938

Hatemi, P. K., Gillespie, N. A., Eaves, L. J., Maher, B. S., Webb, B. T., Heath, A. C., Medland, S. E., Smyth, D. C., Beeby, H. N., Gordon, S. D., Montgomery, G. W., Zhu, G., Byrne, E. M., & Martin, N. G. (2011). A genome-wide analysis of liberal and conservative political attitudes. *Journal of Politics, 73*(1), 1-15. doi:10.1017/S0022381610001015

- Heathers, J. A. (2014). Everything Hertz: methodological issues in short-term frequency-domain HRV. *Frontiers in Physiology*, *5*, 177. doi:10.3389/fphys.2014.00177
- Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences*, *37*(3), 297-350. doi:10.1017/S0140525X13001192
- Hodson, G., Hogg, S. M., & MacInnis, C. C. (2009). The role of "dark personalities" (narcissism, Machiavellianism, psychopathy), Big Five personality factors, and ideology in explaining prejudice. *Journal of Research in Personality*, *43*(4), 686-690. doi:10.1016/j.jrp.2009.02.005
- Huang, L. L., & Liu, J. H. (2005). Personality and social structural implications of the situational priming of social dominance orientation. *Personality and Individual Differences*, *38*(2), 267-276. doi:10.1016/j.paid.2004.04.006
- Iorfino, F., Alvares, G. A., Guastella, A. J., & Quintana, D. S. (2015). Cold face test-induced increases in heart rate variability are abolished by engagement in a social cognition task. *Journal of Psychophysiology*, *30*(1), 38-46. doi:10.1027/0269-8803/a000152
- Jost, J. T., Glaser, J., Kruglanski, A. W. & Sulloway, F. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339-375. doi:10.1037/0033-2909.129.3.339
- Jugert, P., & Hiemisch, A. (2005). Terrorismusbedrohung und sozial-kognitive Motive als Prädiktoren von konservativen Einstellungen. *Zeitschrift für Politische Psychologie*, *13*, 137-158.
- Kanai, R., Feilden, T., Firth, C. & Rees, G. (2011). Political orientations are correlated with brain structure in young adults. *Current Biology*, *21*(8), 677-680. doi:10.1016/j.cub.2011.03.017

- Keltner, D., Kogan, A., Piff, P. K., & Saturn, S. R. (2014). The sociocultural appraisals, values, and emotions (SAVE) framework of prosociality: core processes from gene to meme. *Annual Review of Psychology*, *65*, 425-460. doi:10.1146/annurev-psych-010213-115054
- Kok, B. E., & Fredrickson, B. L. (2010). Upward spirals of the heart: Autonomic flexibility, as indexed by vagal tone, reciprocally and prospectively predicts positive emotions and social connectedness. *Biological Psychology*, *85*(3), 432-436. doi:10.1016/j.biopsycho.2010.09.005
- Kossowska, M., & Van Hiel, A. (2003). The Relationship Between Need for Closure and Conservative Beliefs in Western and Eastern Europe. *Political Psychology*, *24*(3), 501-518. doi:10.1111/0162-895X.00338
- Kteily, N. S., Sidanius, J., & Levin, S. (2011). Social dominance orientation: Cause or “mere effect”? Evidence for SDO as a causal predictor of prejudice and discrimination against ethnic and racial outgroups. *Journal of Experimental Social Psychology*, *47*(1), 208-214. doi:10.1016/j.jesp.2010.09.009
- Kugler, M., Jost, J. T., & Noorbaloochi, S. (2014). Another look at Moral Foundations Theory: Do Authoritarianism and Social Dominance Orientation Explain Liberal-Conservative differences in “Moral” Intuitions? *Social Justice Research*, *27*(4), 413-431. doi:10.1007/s11211-014-0223-5
- Lavine, H., Lodge, M., & Freitas, K. (2005). Threat, Authoritarianism, and Selective Exposure to Information. *Political Psychology*, *26*(2), 219-244. doi:10.1111/j.1467-9221.2005.00416.x
- Lavine, H., Sweeney, D., & Wagne S. H. (1999). Depicting Women as Sex Objects in Television Advertising: Effects on Body Dissatisfaction. *Personality and Social Psychology Bulletin*, *25*(8), 1049-1058. doi:10.1177/0146167299251101

- Lehrer, P. M., & Gevirtz R. (2014). Heart rate variability biofeedback: how and why does it work? *Frontiers in Psychology*, *5*, 756. doi:10.3389/fpsyg.2014.00756
- Ludeke, S. G., & Krueger, R. F. (2013). Authoritarianism as a personality trait: Evidence from a longitudinal behavior genetic study. *Personality and Individual Differences*, *55*(5), 480-484. doi:10.1016/j.paid.2013.04.015
- McCann, S. J. H. (1997). Threatening Times, 'Strong' Presidential Popular Vote Winners, and the Victory Margin, 1824-1964. *Journal of Personality and Social Psychology*, *73*(1), 160-170. doi:10.1037/0022-3514.73.1.160
- McFarland, S. G. (2005). On the eve of war: Authoritarianism, social dominance and American students' attitudes toward attacking Iraq. *Personality and Social Psychology Bulletin*, *31*(3), 360-367. doi:10.1177/0146167204271596
- Milojev, P., Osborne, D., Greaves, L. M., Bulbulia, J., Wilson, M. S., Davies, C. L., et al. (2014). Right-wing authoritarianism and social dominance orientation predict different moral signatures. *Social Justice Research*, *27*(2), 149-174.
- Montano, N., Porta, A., & Malliani, M. (2001). Evidence for Central Organization of Cardiovascular Rhythms. *Annals of the New York Academy of Sciences*, *940*(1), 299-306. doi:10.1111/j.1749-6632.2001.tb03685.x
- Morrison, K. R., & Ybarra, O. (2008). The effects of realistic threat and group identification on social dominance orientation. *Journal of Experimental Social Psychology*, *44*(1), 156-163. doi:10.1016/j.jesp.2006.12.006
- Movius, H. L., & Allen, J. J. B. (2005). Cardiac vagal tone, defensiveness, and motivational style. *Biological Psychology*, *68*(2), 147-162. doi:10.1016/j.biopsycho.2004.03.019
- Negrao, B. L., Bipath, P., Van der Westhuizen, D., & Viljoen, M. (2011). Autonomic correlates at rest and during evoked attention in children with attention-

- deficit/hyperactivity disorder and effects of methylphenidate. *Neuropsychobiology*, 63(2), 82-91. doi:10.1159/000317548
- Otzenberger, H., Gronfier, C., Simon, C., Charloux, A., Ehrhart, J., Piquard, F., & Brandenberger, G. (1998). Dynamic heart rate variability: a tool for exploring sympathovagal balance continuously during sleep in men. *The American Journal of Physiology*, 275(3 pt 2), H946-H950. Retrieved from <http://ajpheart.physiology.org/content/275/3/H946.long>
- Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller, J. L., Scalora, M., Hatemi, P. K., & Hibbing, J. R. (2008). Political attitudes vary with physiological traits. *Science*, 321(5896), 1667-1670. doi:10.1126/science.1157627
- Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., et al. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research*, 59(2), 178-193. doi:10.1161/01.RES.59.2.178
- Park, G., & Thayer, J. F. (2014). From the heart to the mind: cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Frontiers in Psychology*, 5, 278. doi:10.3389/fpsyg.2014.00278
- Park, G., Vasey, M., Van Bavel, J. J., & Thayer, J. F. (2013). Cardiac vagal tone is correlated with selective attention to neutral distractors under load. *Psychophysiology*, 50(4), 398-406. doi:10.1111/psyp.12029
- Porges, S. W. (1992). Vagal tone: A physiological marker of stress vulnerability. *Pediatrics*, 90(3), 498-504. Retrieved from http://www.researchgate.net/publication/21641337_Vagal_tone_a_physiologic_marker_of_stress_vulnerability

- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage: A Polyvagal Theory. *Psychophysiology*, *32*(4), 301-318. doi:10.1111/j.1469-8986.1995.tb01213.x
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, *74*(2), 116-143. doi:10.1016/j.biopsycho.2006.06.009
- Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, *17*(1), 271-320. doi:10.1080/10463280601055772
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741-763. doi:10.1037/0022-3514.67.4.741
- Quintana, D. S., & Heathers, J. A. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology*, *5*, 805. doi:10.3389/fpsyg.2014.00805
- Rickert, E. J. (1998). Authoritarianism and economic threat: Implications for political behavior. *Political Psychology*, *19*(4), 707-720. doi:10.1111/0162-895X.00128
- Rohan, M. J., & Zanna, M. P. (1996). Value transmission in families. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.), *The Ontario symposium: The psychology of values* (Vol. 8, pp. 253-276). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Rokeach, M. (1954). The nature and meaning of dogmatism. *Psychological Review*, *61*(3), 194-204. doi:10.1037/h0060752
- Sales, S. M. (1972). Economic Threat as a Determinant of Conversion Rates in Authoritarian and Nonauthoritarian Churches. *Journal of Personality and Social Psychology*, *23*(3), 420-428. doi:10.1037/h0033157

- Sales, S. M., & Friend, K. E. (1973). Success and Failure as Determinants of Level of Authoritarianism. *Behavioral Science, 18*(3), 163-172. doi:10.1002/bs.3830180304
- Schmitt, M. T., Branscombe, N. R., & Kappen, D. (2003). Attitudes toward group-based inequality: social dominance or social identity? *British Journal of Social Psychology, 42*(2), 161-186. doi:10.1348/014466603322127166
- Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review, 12*(3), 248-279. doi:10.1177/1088868308319226
- Sidanius, J., & Pratto, F. (1999). *Social dominance: an intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.
- Stellar, J. E., Cohen, A., Oveis, C. & Keltner, D. (2015). Affective and physiological responses to the suffering of others: Compassion and vagal activity. *Journal of Personality and Social Psychology, 108*(4), 572-585. doi:10.1037/pspi0000010
- Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–46). Hillsdale, NJ: Lawrence Erlbaum.
- Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability. Standards of measurement, physiological interpretation and clinical use. *Circulation, 93*(5), 1043-1065. Retrieved from http://www.escardio.org/static_file/Escardio/Guidelines/Scientific-Statements/guidelines-Heart-Rate-Variability-FT-1996.pdf
- Thayer, J. F., Ahs F., Fredrikson, M., Sollers J. J., III, & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews, 36*(2), 747-756. doi:10.1016/j.neubiorev.2011.11.009

- Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, 37(2), 141-153. doi:10.1007/s12160-009-9101-z
- Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, 61(3), 201-216. doi:10.1016/S0165-0327(00)00338-4
- Thayer, J. F., & Lane, R. D. (2007). The role of vagal function in the risk for cardiovascular disease and mortality. *Biological Psychology*, 74(2), 224-242. doi:10.1016/j.biopsycho.2005.11.013
- Tritt, S. M., Peterson, J. B., & Inzlicht, M. (2013). Preliminary support for an arousal model of political conservatism. *Plos One*, 8(12), e83333. doi:10.1371/journal.pone.0083333
- Van Hiel, A., Pandelaere, M., & Duriez, B. (2004). The Impact of Need for Closure on Conservative Beliefs and Racism: Differential Mediation by Authoritarian Submission and Authoritarian Dominance. *Personality and Social Psychology Bulletin*, 30(7), 824-837. doi:10.1177/0146167204264333
- Vigil, J. M. (2010). Political leanings vary with facial expression processing and psychosocial. *Group Processes and Intergroup Relations*, 13(5), 547-558. doi:10.1177/1368430209356930gpir.sagepub.com
- Weissflog, M. J., van Noordt, S. J. R., Choma, B. L., Dywan, J., & Segalowitz, S. J. (2010). Sociopolitical ideology and electrocortical responses. *Psychophysiology*, 47(Sup), S24.
- Wendt, A. (1999). *Social Theory of International Politics*. Cambridge: Cambridge University Press.
- Wilson, G. D. (1973). *The psychology of conservatism*. London: Academic Press.

Tonus vagal et attitudes politiques

Les résultats de l'étude 2 montrent qu'un moindre tonus vagal (*i.e.*, moindre RMSSD et HF-HRV) est associé à des scores supérieurs aux échelles RWA et SDO. Nous n'avions pas fait l'hypothèse d'une relation entre tonus vagal et SDO, mais ce résultat est cohérent avec des récentes études montrant une relation entre « menace réaliste » (*i.e.*, menace pour le statut et les ressources du groupe d'appartenance) et SDO (e.g., Morrison & Ybarra, 2008). Selon ces auteurs, l'augmentation du niveau de SDO en réponse à la menace réaliste représente une stratégie cognitive par laquelle les individus se préparent à protéger le statut et les ressources de leur groupe d'appartenance. La théorie polyvagale et le modèle d'intégration neuroviscérale proposent que le tonus vagal soit un biomarqueur de la perception de la menace et du stress. Suivant la perspective polyvagale, les résultats de notre étude suggèrent que la rigidité psychophysiologique représenterait une composante molaire des attitudes politiques. Une surestimation chronique de la menace associée à une capacité réduite de moduler l'influence parasympathique en synchronie avec les demandes externes pourraient sous-tendre la motivation à en endosser des croyances autoritaires. Le modèle d'intégration neuroviscérale conduit à une conclusion similaire : les attitudes politiques refléteraient des différences interindividuelles dans la perception de la menace associée à une dysrégulation de l'intégration neurocardiaque permettant la régulation du métabolisme. En d'autres termes, une difficulté chronique à inhiber le mode de fonctionnement par défaut orienté vers la menace motiverait les individus à adopter des croyances autoritaires et des attitudes punitives.

Ces résultats vont dans le sens de notre hypothèse de travail voulant que l'OD soit la conséquence de stratégies d'autorégulation basées sur l'évitement adoptées dans le but de réduire la détresse provoquée par la résonance empathique. Les deux études de cette partie ont porté sur la relation entre autorégulation et OD.

PARTIE 4. OBEISSANCE DESTRUCTRICE, AUTOREGULATION ET « FATIGUE MENTALE »

4.1. Obéissance destructrice et Tonus vagal

Objectifs et présentation de l'étude

Pour l'étude 3, nous avons exploré deux aspects : (i) l'influence sur l'OD de la flexibilité autonome des participants, (ii) l'exercice d'un contrôle inhibiteur sur les réponses internes chez les participants obéissants. Relativement au premier aspect, nous nous attendions à ce qu'un moindre tonus vagal et une moindre réactivité vagale prédisent l'OD. En ce qui concerne le second aspect, nous nous attendions à retrouver chez les participants obéissants les indices physiologiques d'un effort cognitif (*i.e.*, augmentation de la HRV dans le domaine temporel relativement à la baseline) et une diminution des réponses internes (*i.e.*, diminution de la fréquence cardiaque relativement à la baseline). Nous nous attendions également à ce que la diminution des réponses internes médiatise les effets du tonus vagal, de la réactivité vagale et du RWA sur l'OD. Afin de tester ces différentes hypothèses, nous avons élaboré et utilisé une version alternative de l'IVMOE. Préalablement à leur participation à l'étude, les participants ($n = 85$) ont répondu à un questionnaire en ligne comprenant une version française de l'échelle RWA. Le tonus vagal et la fréquence cardiaque des participants ont d'abord été enregistrées à l'état de repos (5 minutes) puis en continu durant la procédure d'obéissance.

Manuscrit 3⁸

⁸ Manuscrit soumis pour publication auprès de la revue *International Journal of Psychophysiology*.

Autonomic Flexibility and Destructive Obedience

Johan Lepage¹, Laurent Bègue¹, Michael Dambrun², Pierre Chausse², & Martial Mermillod^{3,4}

¹ Laboratoire InterUniversitaire de Psychologie, Université Grenoble Alpes, Grenoble, France

² Laboratoire de Psychologie Sociale et Cognitive, CNRS UMR 6024, Université Blaise Pascal, Clermont-Ferrand, France

³ Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105, Université Grenoble Alpes, Grenoble, France

⁴ Institut Universitaire de France, Paris, France

Corresponding author

Johan Lepage, PhD Candidate: Laboratoire InterUniversitaire de Psychologie, UFR SHS, 1251 avenue Centrale, BP 47, 38040 Grenoble Cedex 9, France; E-mail: johan.lepage@gmail.com, Phone: (+33 4) 76 82 58 92, Fax: (+33 4) 76 82 56 65

Abstract

The present study tested the hypothesis that individuals with greater autonomic flexibility, operationalized as higher cardiac vagal tone at rest and greater vagal reactivity (indexed by a decrease in high frequency heart rate variability during challenge), would be more able to disobey destructive orders in a Milgram-like paradigm. We also hypothesized that emotion regulation strategy of avoidance (operationalized in term of inhibitory control over automatic emotional response) would mediate the influence of autonomic flexibility on destructive obedience. Heart rate variability was collected during 5-min baseline and continuously during the obedience procedure. Prior to the experiment, the participants ($n = 85$) also completed a survey assessing right-wing authoritarianism (RWA), a well-known predictor of destructive obedience. Using the Immersive Video Milgram Obedience Experiment, we found that lower vagal tone at rest and lower vagal reactivity predicted higher destructive obedience. We also found some evidence that obedient participants have exerted cognitive effort to inhibit their automatic emotional responses to the victim's pain (i.e., increased RMSSD and decreased heart rate relative to baseline). Inhibitory control over automatic response (indexed by a decrease in heart rate relative to baseline) mediated the influence of vagal tone, vagal reactivity and RWA on destructive obedience. Implications for future directions in obedience research are discussed.

Keywords

Obedience to authority, right-wing authoritarianism, autonomic flexibility, inhibitory control

Introduction

It is now more than 50 years since Milgram (1963) ignited the experimental study of destructive obedience (i.e., obedience to orders that cause harm to another individual). In Milgram's obedience experiments, naive participants were ordered to administer increasingly severe electric shocks (ranging from 15 to 450 volts) on a "learner" (a confederate) after being told that they were participating in an experiment on the effects of punishment on learning. In the so-called "baseline" condition (condition 5; Milgram, 1974), results revealed that 65% of the participants were willing to administer allegedly lethal electric shocks when ordered to do so. Few studies aimed at "capturing" individual differences in reactions to destructive orders and indicate that authoritarianism can predict destructive obedience. Adorno, Frenkel-Brunswik, Levinson and Sanford (1950) first defined authoritarianism as a "personality syndrome" which they attempted to measure with their F-scale (F for fascist). Elms and Milgram (1966) pooled a subsample of participants from the first four Milgram's experiments and found that fully obedient participants scored higher on the F-scale. The Altemeyer's right-wing authoritarianism (RWA) scale (Altemeyer, 1981, 1988) incorporated a reconceptualization of authoritarianism as well as important psychometric refinements. The items capture the covariation of three dimensions: (i) authoritarian submission (a submissive, uncritical attitudes toward authorities), (ii) authoritarian aggression (tendency to condemn, reject, and punish people who violate conventional values), and (iii) conventionalism (rigid adherence to conventional middle-class values). Studies found that participants scoring higher on the RWA scale were those who administered more electric shocks (e.g., Altemeyer, 1981; Dambrun & Vatiné, 2010).

Despite the passage of time, Milgram's experiments are still widely cited when explaining destructive obedience (e.g., Kelman & Hamilton, 1989; Lankford, 2009; Waller,

2002). These experiments have also been embroiled in an intense ethical controversy (particularly because of the deception involved; e.g., Miller, 1986) and current ethical standards place Milgram's studies out of bounds (e.g., Elms, 1995). As a result, very few experimental studies have been conducted since the pioneering Milgram's work and the mechanisms responsible for destructive obedience remain unknown (e.g., Blass, 1991; Cheetham, Pedroni, Antley, Slater, & Jäncke, 2009). However, experimental research was recently reopened through the use of immersive environments and new data point out the role of self-regulation processes.

Affective-motivational responses to the others' pain

Milgram (1963, 1974) emphasized that many participants exhibited signs of emotional distress. However, whether the participants' response to the victim's pain was more akin to other-oriented empathic concern or to a self-oriented aversive state of personal distress is unclear (Cheetham et al., 2009). Research exploring how people respond to the others' pain demonstrates that pain affective sharing can provoke either sympathy (an other-oriented motivated response) or personal distress (an aversive, self-oriented motivated response; Decety & Lamm, 2009). While sympathy initiates care-based moral reasoning and prosocial behavior (i.e., actions that are intended to help or benefit another individual; Decety & Svetlova, 2012), personal distress is associated with the desire to alleviate one's own but not the other's distress (e.g., Eisenberg & Eggum, 2009). Self-regulation processes are believed to determine the affective-motivational responses to the others' pain (e.g., Decety & Lamm, 2009; Eisenberg & Eggum, 2009). Consistently, studies found that "empathic overarousal" promotes a self-focus and a desire to alleviate one's own, but not the other's, negative arousal (e.g., Eisenberg, Valiente, & Champion, 2004). Studies also found that empathy-related

responses are malleable with respect to several factors such as contextual appraisal (e.g., de Vignemont & Singer, 2006).

The Milgram paradigm induces a social dilemma in which participants are wedged between conflicting alternatives “for which there is no prospect of establishing a personally and collectively favorable outcome” (Cheetham et al., 2009). Using a virtual version of the Milgram paradigm, the authors found that the observation of the victim’s pain within this context induced brain activations related to an aversive state of personal distress (i.e., activations in pre-defined brain areas involved in affective processing - e.g., right amygdala - but not in those commonly associated with emotion regulation - e.g., anterior cingulate cortex). These findings suggest that low self-regulatory control of the shared affect evoked by the emotional state of the victim could be responsible (at least partially) for destructive obedience.

Given these results, cardiac vagal tone (VT) seems to qualify as a potential predictor of destructive obedience. A plethora of studies demonstrates that VT reflects self-regulation processes and represents a robust biomarker of stress vulnerability (e.g., Porges, 2007).

Autonomic flexibility

The autonomic nervous system (ANS) is subdivided into an excitatory sympathetic nervous system (SNS) and an inhibitory parasympathetic nervous system (PNS). When an individual is threatened, the activity of the SNS becomes dominant, triggering a suite of physiological responses to aid in adapting to the challenge (e.g., increased heart rate). In a safe context, the PNS is dominant and maintains a lower degree of physiological arousal (e.g., decreased heart rate). A flexible ANS allows for rapid modulation of physiological states in accordance with situational demands. In contrast, autonomic rigidity results in a lessened capacity to alter physiological responses in synchrony with environmental changes (e.g.,

Appelhans & Luecken, 2006). The polyvagal theory (e.g., Porges, 1995, 2007) causally relates autonomic flexibility with emotion regulation and social behavior.

Polyvagal theory

The polyvagal theory (e.g., Porges, 1995, 2007) articulates three phylogenetically-ordered neural circuits regulating the heart: (i) the “social engagement system” is dependent upon the functions of the myelinated vagus “which serves to foster calm behavioral states by inhibiting the sympathetic influences to the heart” (Porges, 2007); (ii) the “fight-or-flight system” is dependent on the functioning of the SNS; (iii) the “immobilization system” (the most phylogenetically primitive component) is dependent on the unmyelinated vagus. In safe contexts, there is an increased influence of the myelinated vagus on the cardiac pacemaker that slows the heart and inhibits the fight/flight mechanisms of the SNS. The two more primitive neural circuits serve to ensure survival in life-threatening contexts through defensive strategies (i.e., fight–flight and death-feigning behaviors). Under resting conditions, the myelinated vagus acts as a “brake”, firing a rapid and continuous signal that slows the heart.

The polyvagal theory also states that a neural process termed “neuroception” continuously evaluates risk and modulates vagal activity via higher brain structures. An invalid neuroception of danger is believed to contribute to maladaptive physiological reactivity and the expression of the defensive behaviors associated with negative emotional states (e.g., fear, anxiety; Porges, 2007).

The strength of the tonic vagal influence is commonly assessed by examining resting heart rate variability (HRV). Following the polyvagal perspective, higher resting HRV should be related to higher self-regulatory control and prosocial behaviors. Consistently, studies found that higher resting HRV predicted greater regulatory control, decreased negative

emotional arousal in response to stressors, sympathy, compassion and helping behavior (e.g., Fabes & Eisenberg, 1997; Stellar, Cohen, Oveis, & Keltner, 2015). By contrast, lower resting HRV has been related to personal distress, psychosocial stress, social anxiety and defensiveness (e.g., Movius & Allen, 2005; Porges, 1992; Thayer & Lane, 2007). More generally, a “host of findings suggest that human subjects with greater vagal influence may be dispositionally more open to experience and better able to self-regulate” (Movius & Allen, 2005).

Model of neurovisceral integration

The model of neurovisceral integration (e.g., Thayer & Lane, 2000; Thayer et al., 2009) states that the neural networks implicated in cognitive and emotion self-regulation (e.g., ventromedial prefrontal cortices, anterior cingulate, insula, amygdala) are also involved in the control of cardiac activity. In safe contexts, the prefrontal cortex is believed to exert an inhibitory control over sympatho-excitatory subcortical circuits (e.g., Thayer et al., 2009). In threatening contexts, the prefrontal inhibitory regulation diminishes and sympatho-excitatory subcortical circuits produce default threat responses (e.g., Park, Vasey, Van Bavel, & Thayer, 2013). Consistently, brain imaging studies found that reduced prefrontal activity is responsible for hyperactive subcortical activity, prolonged activation of defensive behaviors and hyper-vigilance (e.g., Thayer et al., 2009; Park et al., 2013). Research also provides evidence that the functioning of prefrontal-subcortical inhibitory circuits is linked with the heart via the vagus nerve. For instance, studies found that lower VT at rest is associated with hypoactive prefrontal regulation, hyperactive subcortical structures, maladaptive cognitive and emotion regulation, and increased negativity bias (i.e., the tendency to prioritize negative information over positive; see Thayer, Ahs, Fredrikson, Sollers, & Wager, 2012). In this

perspective, VT reflects inhibitory prefrontal processes and represents a useful index of threat perceptions and therefore of stress.

VT is usually measured as HRV recorded during 5 min at rest, but few studies have examined vagal reactivity (VR), or decreases in high frequency HRV in response to challenge. The high frequencies (HF-HRV, 0.15-0.4 Hz) are thought to reflect greater myelinated vagal influence, and as such greater parasympathetic nervous system influence on the heart (e.g., Montano, Porta, & Malliani, 2001). Previous research suggests (i) that cognitive, social, emotional, and sensory challenge tasks elicit vagal withdrawal, or decreases in HF-HRV from baseline (e.g., Van Roon, Mulder, Althaus, & Mulder, 2004), (ii) that the magnitude of HF-HRV decrease is associated with better task performance and greater prosociality (e.g., Akinola & Mendes, 2014; Obradović, Bush, Stamperdahl, Adler, & Boyce, 2010). A recent study has examined the association between VR and social sensitivity (Muhtadie, Koslov, Akinola, & Mendes, 2015). The authors found that greater VR was related to greater social sensitivity and responsiveness to static facial images and dynamic social interactions. They also found that in the context of social rejection cues, greater VR was associated with greater negative responses, including more shame, higher blood pressure and less sociable behavior (trained coders rated participants on several aspects of behavior that occurred during the interaction, including “smiling and laughing” and “gesturing and animated”). In the context of social acceptance cues, greater VR translated into lower blood pressure and more sociable behavior.

It is reasonable to hypothesize that higher autonomic flexibility (i.e. higher VT at rest and higher VR) would promote disobedience toward destructive orders within the Milgram paradigm. Recent research also suggests that emotion regulation strategy of avoidance could mediate the influence of autonomic flexibility on social behavior.

Emotion down-regulation processes

There is a growing appreciation that distressed individuals use a wide range of strategies for down-regulating aversive emotions (Gross, 2002; Gross & John, 2003). Emotion regulation strategy of avoidance (e.g., escape behaviors, inhibition of internal responses, cognitive reappraisal) is a common maladaptive response to aversive emotional experiences (e.g., Aldao, Dixon-Gordon & De Los Reyes, 2015). According to Goubert, Craig and Buysse (2009), individuals responding with high distress to the other's pain could be motivated "to underestimate the observed person's pain, in an attempt to keep their own distress within acceptable limits". As a consequence, individuals may tend not to fully attend to the other's experience and fail to display sympathetic responses. Brain imaging studies found a decreased amygdala activity during the observation and the experience of pain that indicate a cognitive strategy to attenuate pain-related stress responses in a "no-escape situation" (e.g., Petrovic, Carlsson, Petersson, Hansson, & Ingvar, 2004). Because decreased activity within the amygdala has been linked to decreased harm aversion (e.g., Blair, 2007; Moll & de Oliveira-Souza, 2007), such an inhibitory control could have detrimental consequences in a Milgram-like paradigm (i.e., increased destructive obedience). Aldao et al. (2015) found evidence that higher VT at rest might protect against the use of avoidance in response to aversive stimuli.

Research on emotion regulation suggests that low autonomic flexibility would favor destructive obedience *via* inhibitory control over automatic emotional responses. Interestingly, research aiming to explain the origins and manifestations of political attitudes suggests that low autonomic flexibility and inhibitory control may be responsible for the well-known influence of RWA on destructive obedience. Indeed, a long tradition in political psychology research posits that authoritarianism is related to cognitive rigidity, threat sensitivity and defensiveness (e.g., Duckitt, 2013; Feldman, Lavine, Lodge & Verhulst, 2010;

Weissflog et al., 2010). Following a polyvagal perspective, one could hypothesize that autonomic rigidity is a core feature by which we may understand RWA. That is, RWA may reflect individuals differences in a predisposition to chronic threat perception associated with dysregulated brain-peripheral integration.

Hypotheses

First, we hypothesized that lower VT at rest and lower VR would predict higher RWA scores and destructive obedience. *Second*, we hypothesized that obedient individuals would exert an inhibitory control over automatic responses to the victim's pain and distress in an attempt to reduce their own negative arousal. *Third*, we hypothesized that inhibitory control would mediate the influence of VT, VR and RWA on destructive obedience. Previous research demonstrates that cognitive effort increases HRV (measured as the root mean squared successive differences in the interbeat interval – RMSSD; e.g., Segerstrom & Ness, 2007) and that the SNS responses to aversive stimuli are associated with increased heart rate (e.g., Cacioppo, Tassinary, & Bernston, 2007; Eisenberg, Fabes, & Spinrad, 2006). Consistently with the idea that destructive obedience requires inhibitory control over automatic responses, we also predicted increased RMSSD and decreased heart rate relative to baseline among obedient participants.

Method

Participants

Eighty-five healthy adults (51 females, $M_{age} = 25.17$, $SD = 5.43$) volunteered for the experiment in exchange for 30€. Participants were French or bilingual in French. They provided written informed consent before participation. The study was part of a wider research project on the role of self-regulation processes in the Milgram paradigm, reviewed

and approved by the University human ethics committee from Grenoble, France (Grenoble ethics committee notice number 2013-07-04-23). The protocol was carried out in accordance with the Helsinki Declaration, Convention of the Council of Europe on Human Rights and Biomedicine.

To be eligible, participants had to be aged between 18 and 50 years, with a normal or normal-to-corrected vision, explicitly reported an absence of psychiatric, neurologic, hormonal, or cardio-vascular disease, and with no medical treatment (with the exception of contraception). Individuals who might be familiar with obedience research were not included (participants were asked if they were familiar with several famous social psychology experiments, including Milgram's studies and "Le Jeu de la Mort", a transposition of the Milgram paradigm to the context of a TV game show, broadcasted in 2010; Beauvois, Courbet, & Oberlé, 2012). Smoking, energizing drinks (e.g., coffee, tea, etc.) and psychotropic substances (e.g., alcohol, cannabis, etc.) were prohibited to each participant the day of the experiment. They had also to avoid eating or drinking (water was allowed) during the 2 hours preceding the experiment in order to limit the influence of digestion on autonomic functioning but they had to eat in the morning (more than 2 hours before the experiment) in order to avoid fasting states (short term HRV measurement can be biased by the digestion of food since viscera are innervated by the autonomic nervous system; Heathers, 2014; Iorfino, Alvares, Guastella, & Quintana, 2015; Quintana & Heathers, 2014). Finally, they had to avoid water consumption 30 min before the experiment and were asked to empty their bladder before starting the experiment.

R-to-R measurement

Measurements of HRV include time domain and frequency domain methods. In the former domain, the most popular estimate of HRV is RMSSD (root mean square of

successive R-R differences), which is calculated as the standard deviation of the beat-to-beat intervals. In the latter domain, a popular technique to estimate HRV involves decomposing heart period variance into different frequency bands using Fourier transformations. As mentioned above, HF-HRV (0.15-0.4 Hz) are thought to represent primarily vagal influence, and as such parasympathetic activity (e.g., Montano , Porta, & Malliani, 2001). Studies found that RMSSD and HF-HRV are positively correlated (e.g., Otzenberger et al., 1998). The low frequencies (LF-HRV, 0.04-0.15 Hz) are believed to represent the influence of both sympathetic and parasympathetic activities (see Negrao, Bipath, van der Westhuizen, & Viljoen, 2011, for a summary).

Resting HRV was recorded during 10 min. Three Ag-AgCl electrodes were attached as follows: one electrode (Negative) on right clavicle and one electrode (Positive) on the left 4th or 5th rib mid-clavicular to axillary line, with the Ground on right 4th or 5th rib mid-clavicular to axillary line. ECG (sampling rate 1000 Hz) was recorded using Single BIO Amp hardware in conjunction with PowerLab 16/32 hardware and LabChart 8 software (ADInstruments Ltd, UK). HRV was also recorded continuously during the obedience procedure.

Procedure

Recruitment and questionnaire. Participants were recruited through newspaper advertisements, public notices, flyers and direct email solicitation, to participate in a research on “memory and cardiac function”. Those interested were assessed for eligibility through an online form (see criteria above). Then, eligible participants were asked to fill out an online survey including 10 items ($\alpha = .83$) from the French translation of Altemeyer’s 20-item (1988) RWA Scale (1 = *totally disagree* to 7 = *totally agree*).

HRV baseline. Participants were greeted by the experimenter (30 years-old male in a white technician's coat). The experiment took place in a quiet and dark room; participants were seated on a comfortable chair. They were told that they will participate to two experiments: a first one about cardiac function, and a second one about memory. After being fully informed about the "first experiment", participants gave their written consent for participation. They were told that their cardiac activity will be recorded during 10 min and that they should try to relax. The 5 first min allowed participant to shift in a quiet state and the HRV baseline was recorded during the 5 following min. When this step ended, the experimenter introduced the "second experiment".

Obedience paradigm. We used an alternative version of the Immersive Video Milgram Obedience Experiment (IVMOE; Dambrun & Vatiné, 2010). Recent studies have found immersive virtual environment technology useful for studying social behavior (e.g., Blascovich, Loomis, Beall, Swinth, Hoyt, & Bailenson, 2002), and Slater et al. (2006) first carried out a replication of Milgram's experiment using an immersive virtual environment. While the victim was a virtual character, the results showed that the participants responded to the situation at the behavioral, physiological and subjective levels as if it was real. The IVMOE is very similar in content to the original Milgram paradigm except that the victim (an actor filmed and recorded) is seen onscreen. All parameters are computer-programmed in order to give the impression of a real time and dynamic environment. Dambrun and Vatiné (2010) found that participants responded to the situation at the subjective and behavioral levels as if the victim and the shocks were real (e.g., participants experienced discomfort and expressed the desire to discontinue or did actually stop delivering electric shocks). When the victim was hidden, the level of obedience they obtained was similar to Milgram's. Importantly, the authors replicated several previous findings observed in real environments: (i) participants were more obedient when the victim was hidden (vs. visible), (ii) the more

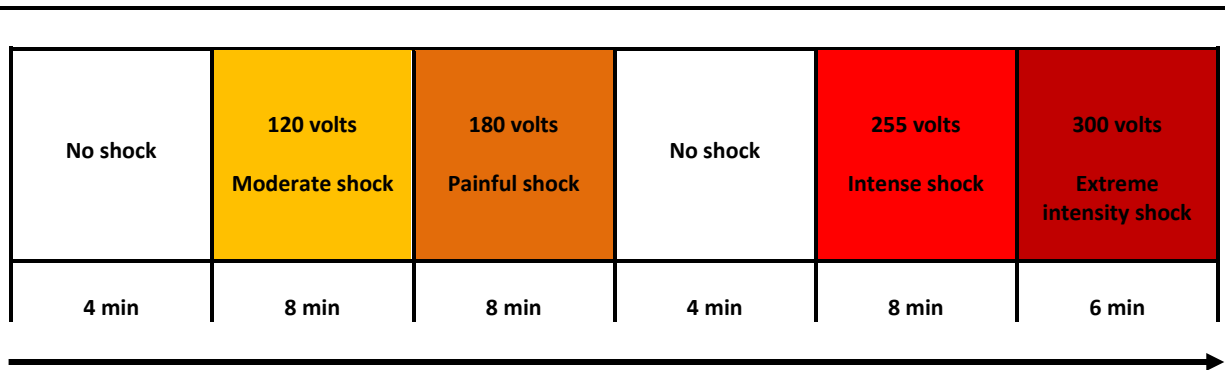
participants negated their own responsibility by projecting responsibility on both the victim and the experimenter, the more they obeyed, (iii) participants scoring higher on the RWA scale were those who administered more violent shocks.

First, participants in our study were told that the experiment was a simulation, as follow: “This experiment will be role-play. You will interact with another person; this is an actor that will simulate reactions to the experimental procedure. We ask you to behave exactly as if everything was real. You will have to make decisions during the experiment; we ask you to make these decisions not by considering that it will have no consequences because of the role-play, but by considering that each of your decision will have concrete consequences on real persons within a real situation. This is the most important point: you have to behave as if everything was real. If you are ok with this, right now I will behave myself as if everything was real. If you ask me about the role play, I will respond that everything is real”.

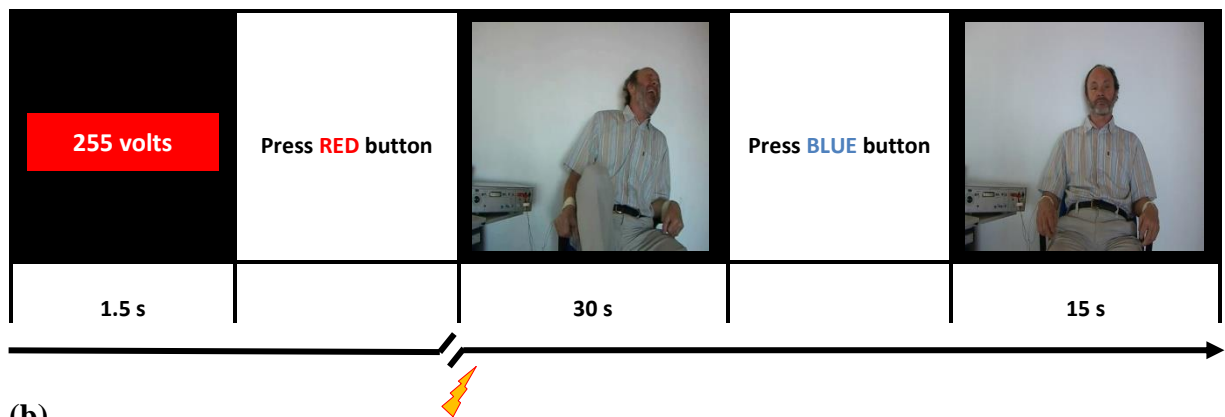
Then, participants were given a booklet about the purpose of the experiment (cover-story) and the experimental procedure. First, participants were informed that they were involved in an experiment on the effects of defibrillation on human memory; it was claimed that very little is known about the short-term effects of defibrillation on memory and how much these effects could bias medical diagnosis (i.e., misattribution of memory impairment). Participants were informed that we were bringing together a number of adults; we were asking some of them to pass a memory test and to receive electric shocks during this test (the “learners”), and some of them to administrate the shocks (the participants). The” learners” and the participants were supposed to be installed into two adjacent rooms. A video-conferencing system was supposed to ensure the communication. Then, participants were informed about the experimental procedure. The memory test was a paired-associate learning task. Before the experiments, the “learners” were supposed to memorize a series of word pairs; during the

experiment, they were to indicate for each pair, which of four words had originally been paired with the first word. It was emphasized that to ensure the safety of the procedure (i) the memory test was administered by a doctor, (ii) the shocks were delivered on the arms (and not on the chest), (iii) “learners” were strapped on a chair to avoid any dangerous movements. The notice provided very simple instructions to the participants: they were asked to follow the instructions appearing on the screen during the experiment (“press red button” to deliver the shock, “press blue button” to go to the next question). The last part provided Figure 1(a) with the following explanations: (i) the experimental procedure was divided into six blocks, (ii) participants were asked to administer a shock before each question into four blocks (see Figure 1). It took approximately 5 min to the participants to read and memorize the booklet content. After having read the booklet, participants were invited to ask any questions about the procedure. Then, they gave their written consent for the participation.

The IVMOE task was adapted from Dambrun & Vatiné, (2010). The procedure started with the seating of the student on a chair by the virtual doctor (35 years-old male in a white technician’s coat – 1.25 min): the learner was strapped into the chair and an electrode was attached on each arm. The next phase was the experimental procedure. In our immersive obedience paradigm, each trial begins by indicating the shock intensity (e.g., 120V) to be administered to the learner (see Figure 1(b)). At the level of 120V, the learner exhibited some cues of pain-related behavior (moan, facial expression of pain and corresponding body movement). At the level 180V, the learner’s pain became more intense. At the level of 255V, the learner reacted strongly and asked to stop the experiment because the shocks were too painful. At the level of 300V, the learner’s groaning was loud and asked to stop the experiment.



(a)



(b)

Figure 1. Experimental procedure. (a) Overview of the experimental procedure. The experimental procedure was structured into six blocks, with five trials (i.e. five questions to the learner) in the first five blocks, and three trials in the last block. The total experimental time was approximately 40 min. (b) Overview of the timeline of a pain trial. Each trial begins by indicating the shock intensity (e.g., 120V) to be administered to the learner. In the pain trials, participants had to press the red button and the subsequent suffering of the learner was observed. Each shock is accompanied by an “electric” buzz sound. Then participants had to press the blue button. The words were enunciated by the virtual doctor and the learner gave his answer. In the non-pain trials, participants had to press the blue button and the learner was simply observed giving his answer (30 s). All parameters were computer-programmed in order to give the impression of real time and a dynamic environment.

When participants expressed their desire to stop the experiment, the experimenter intervened with four consecutive injunctions: (i) “Please continue”, (ii) “It is important for the experiment that you continue”, (iii) “The procedure requires that you continue”, (iv) “You have no other choice, you must go on”. If participants insisted on stopping the experiment after the fourth injunction, the experiment was terminated. Otherwise, this four-part sequence was restarted each time the participants asked whether they could stop the experiment.

Obedience was measured as follow: (i) absolute obedience was indicated by whether participants delivered all the shocks (0 = *not fully obedient*, 1 = *fully obedient*), (ii) the relative obedience refers to the maximum number of shocks administrated (from 0 to 18). Participants were fully debriefed after the experiment about their experience during the obedience procedure with particular reference to the impact of observing the learner in pain.

Data analysis

Prior to analysis, signal was visually inspected and ectopic (abnormal) heart beats were removed using the LabChart Beat Classifier view. The HRV 2.0 Add-On for LabChart was used to calculate heart rate, RMSSD (root mean square of successive R-R differences), HF-HRV (0.15-0.4 Hz; normalized units) and LF-HRV (0.04-0.15 Hz; normalized units). The HRV 2.0 Add-On module uses the Lomb Periodogram nonparametric method for spectral analysis which, unlike the Fast Fourier Transform method, can compute the spectral frequencies directly from the unevenly sampled tachogram. As the total power of the spectral signal is heterogeneous from person to person, it has been recommended that HF-HRV and LF-HRV data should be presented as normalized values (e.g., Pagani et al., 1986).

Our hypotheses were tested using binary logistic regression, simple linear regression, and repeated measures ANOVA. We tested our mediation hypotheses following the procedure advocated by Preacher and Hayes (2004). All our continuous predictors were centered before analysis. We calculated VR by subtracting the mean HF-HRV during the first block “No shock” from the baseline HF-HRV (higher scores indicated greater VR). Inhibitory control was calculated by computing the difference score between heart rate baseline and heart rate procedure (higher scores indicated an increase in inhibitory control).

Results

Rate of obedience

The distribution of break-off points is presented in table 1. On the whole, 67.1% of participants fully obeyed. This is very close to the rates of obedience found by Milgram in his “baseline” condition (65%) and his “vocal feedback” condition (62.5%).

Table 1

Distribution of break-off points

Shock level	Verbal designation and voltage level	Number of participants for whom this was the maximum shock
	Refuse	6
	No shock	
	0	
	0	
	0	
	0	
	0	
	Moderate shock	
1	120	
2	120	3
3	120	
4	120	
5	120	1
	Painful shock	
6	180	
7	180	5
8	180	3
9	180	1
10	180	1
	No shock	
	0	
	0	
	0	
	0	
	0	
	Intense shock	
11	255	1
12	255	2
13	255	1
14	255	1
15	255	1
	Extreme intensity shock	
16	300	2
17	300	
18	300	57
Percentage of obedience		67.1
Mean shock level		14.4

Data description

Table 2 provides descriptive statistics and Pearson correlation coefficients between the variables included in the present study. Figure 2 provides scatter plots and densities.

Table 2

Descriptive statistics and correlations between obedience, right-wing authoritarianism and physiological parameters

	1	2	3	4	5	6	7
1 R. obedience	—	0.23 *	-0.29 **	-0.27 *	0.28 **	-0.26 *	0.52 ***
2 RWA		—	-0.24 *	-0.31 **	0.32 **	-0.29 **	0.22 *
3 RMSSD			—	0.48 ***	-0.49 ***	0.12	-0.44 ***
4 HF-HRV _{n.u.}				—	-0.99 ***	0.61 ***	-0.40 ***
5 LF-HRV _{n.u.}					—	-0.62 ***	0.40 ***
6 VR						—	-0.40 ***
7 IC							—
<i>M</i> (<i>SD</i>)	14.40 (5.98)	3.14 (1.04)	42.06 (26.64)	43.82 (21.93)	56.14 (22.39)	.88 (18.71)	1.35 (8.05)

Note. R. obedience = relative obedience; RWA = right-wing authoritarianism; RMSSD = root mean square of successive R-R differences; HF-HRV_{n.u.} = high frequencies (0.15-0.4 Hz) expressed in normalized units; LF-HRV_{n.u.} = low frequencies (0.04-0.15 Hz) expressed in normalized units; VR = vagal reactivity; IC = inhibitory control. * $p < .05$ ** $p < .01$ *** $p < .001$

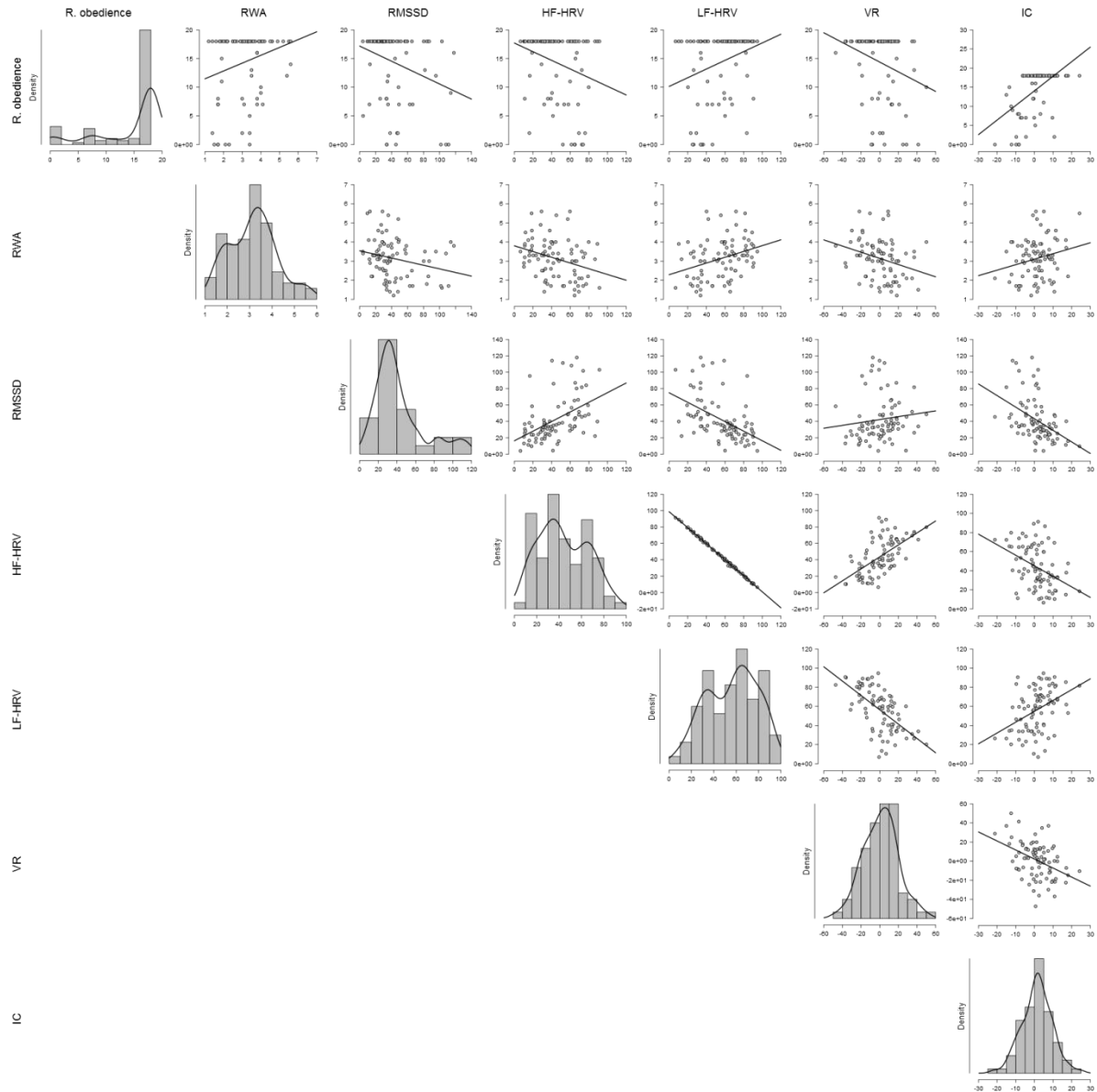


Figure 2. Scatter plots, densities, and Pearson correlation coefficients between recorded variables. R. obedience = relative obedience; RWA = right-wing authoritarianism; RMSSD = root mean square of successive R-R differences; HF-HRV = high frequencies (0.15-0.4 Hz) expressed in normalized units; LF-HRV = low frequencies (0.04-0.15 Hz) expressed in normalized units; VR = vagal reactivity; IC = inhibitory control

Autonomic flexibility and Destructive Obedience

Results are presented in Table 3. Because of multicollinearity, physiological parameters were not entered simultaneously.

Table 3

Regression analyses examining the associations between physiological parameters and obedience

Criterion variable	Predictor variables	β	t	p	95% CI	
Relative obedience	RMSSD	-.29	-2.80	.006	[-3, -.51]	
	HF-HRV _{n.u.}	-.23	-2.18	.03	[-2.66, -.12]	
	LF-HRV _{n.u.}	.23	2.24	.02	[.16, 2.69]	
	VR	-.26	-2.46	.01	[-2.88, -.31]	
Criterion variable	Predictor variables	OR	SE	W	p	95% CI
Absolute obedience	RMSSD	1.61	.25	8.31	.004	[.29, .79]
	HF-HRV _{n.u.}	1.87	.23	3.56	.05	[.38, 1.01]
	LF-HRV _{n.u.}	4.80	.23	3.55	.05	[.98, 2.51]
	VR	1.85	.25	3.57	.05	[.38, 1.01]

Note. RMSSD = root mean square of successive R-R differences; HF-HRV_{n.u.} = high frequencies (0.15-0.4 Hz) expressed in normalized units; LF-HRV_{n.u.} = low frequencies (0.04-0.15 Hz) expressed in normalized units; VR = vagal reactivity.

Table 3 shows that the markers of parasympathetic activity and VR were associated with lower obedience. We also found that higher RMSSD, HF-HRV and VR predicted lower RWA scores ($b = -.24, t = -2.26, p = .02$; $b = -.32, t = -3.07, p = .003$; $b = -.29, t = -2.73, p = .008$), while higher LF predicted higher RWA scores ($b = .33, t = 3.19, p = .002$).

Inhibitory control

RMSSD of fully obedient individuals was higher during the procedure relative to baseline, as reflected in a significant interaction of time (baseline vs. procedure) and obedience, $F(1, 82) = 15.60, p = .000$. We also found that HR of fully obedient individuals was lower during the procedure relative to baseline, as reflected in a significant interaction of time (baseline vs. procedure) and obedience, $F(1, 82) = 34.98, p = .000$ (see Figure 3).

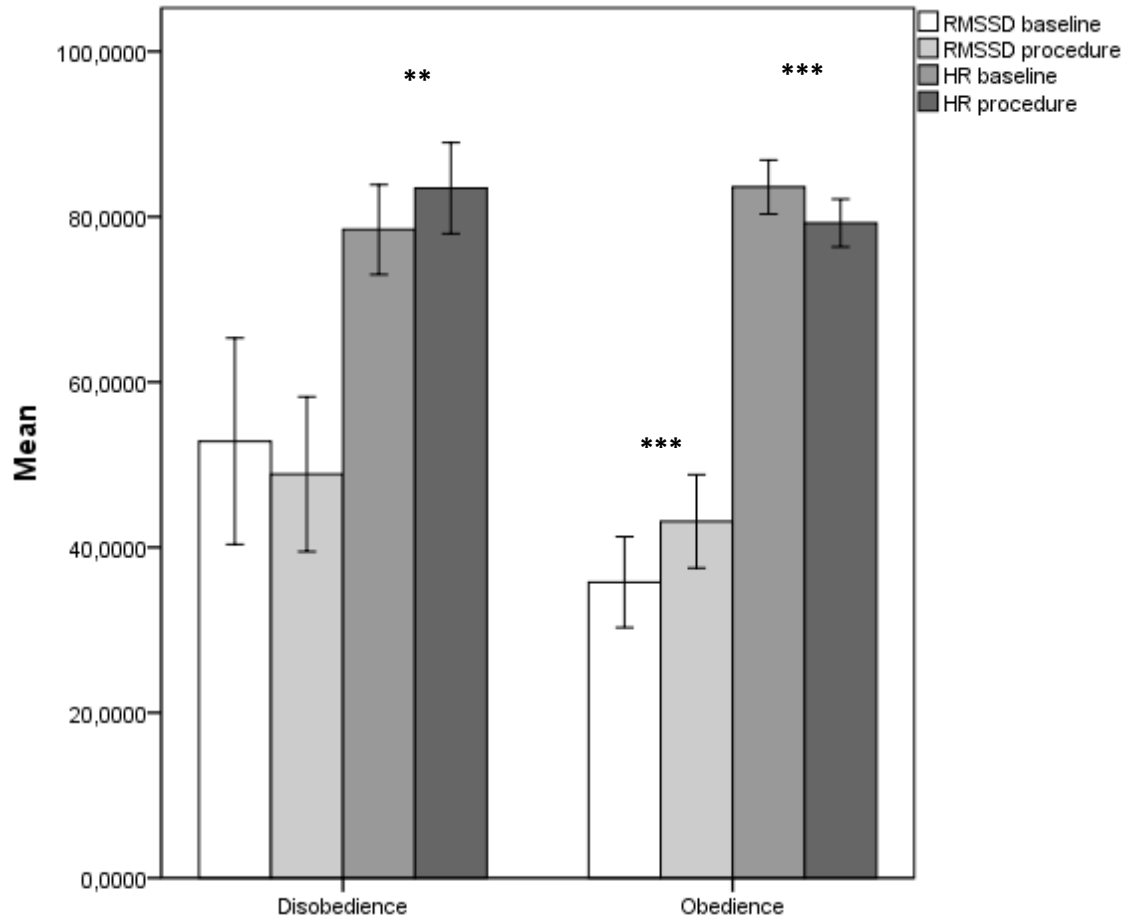


Figure 3. Heart rate variability (RMSSD) and heart rate (beats per minute) changes among obedient and disobedient participants. RMSSD = root mean square of successive R-R differences; HR = heart rate. * $p < .05$ ** $p < .01$ *** $p < .001$

Mediation analyses

We tested the extent to which inhibitory control mediated the influence of VT at rest, VR and RWA on relative obedience using the bootstrapping method advocated by Preacher and Hayes (2004). As mentioned above, regression analyses resulted in a significant effect of VT at rest (i.e., RMSSD and HF-HRV) and VR on relative obedience. We also found a significant effect of RWA on relative obedience, $b = .23$, $t = 2.24$, $p = .02$. In the first step, we used inhibitory control as the criterion variable and found (i) that higher RMSSD predicted lower inhibitory control, $b = -.45$, $t = -4.45$, $p = .000$, (ii) that higher HF-HRV predicted lower inhibitory control, $b = -.40$, $t = -4$, $p = .000$, (iii) that higher VR predicted lower inhibitory control, $b = -.41$, $t = -4$, $p = .000$, (iv) that higher RWA predicted higher inhibitory control, $b = .21$, $t = 2.04$, $p = .04$. The relationship between inhibitory control and relative obedience remained significant even while controlling for (i) RMSSD, $b = 3.05$, $t = 5$, $p = .000$, (ii) HF-HRV, $b = 2.93$, $t = 4.91$, $p = .000$, (iii) VR, $b = 2.84$, $t = 4.71$, $p = .000$, and (iv) RWA, $b = .288$, $t = 5.21$, $p = .000$. The effect of RMSSD on relative obedience controlling for inhibitory control became non-significant, $b = -.04$, $t = -.07$, $p = .94$. Employing the bootstrapping method (with 5000 samples) produced a 95% bias-corrected confidence interval of [-2.39, -.76] for the indirect effect of RMSSD on relative obedience through inhibitory control ($z = -3.29$, $p = .001$). The confidence interval excluded zero, which suggests a significant indirect effect (see Figure 4a). The effect of HF-HRV on relative obedience controlling for inhibitory control became non-significant, $b = -.34$, $t = -.58$, $p = .56$ (95% CI [-2.19, -.52] for the indirect effect of HF-HRV on relative obedience through inhibitory control, $z = -3.06$, $p = .002$; see Figure 4b). The effect of VR on relative obedience controlling for inhibitory control became non-significant, $b = -.51$, $t = -.84$, $p = .40$ (95% CI [-2.08, -.48] for the indirect effect of VR on relative obedience through inhibitory control, $z = -3.01$, $p = .002$; see Figure 4c). Finally, the effect of RWA on relative obedience controlling for inhibitory control became non-

significant, $b = .84$, $t = 1.52$, $p = .13$ (95% CI [.03, 1.56] for the indirect effect of RMSSD baseline on relative obedience through inhibitory control, $z = 1.87$, $p = .06$; see Figure 4d).

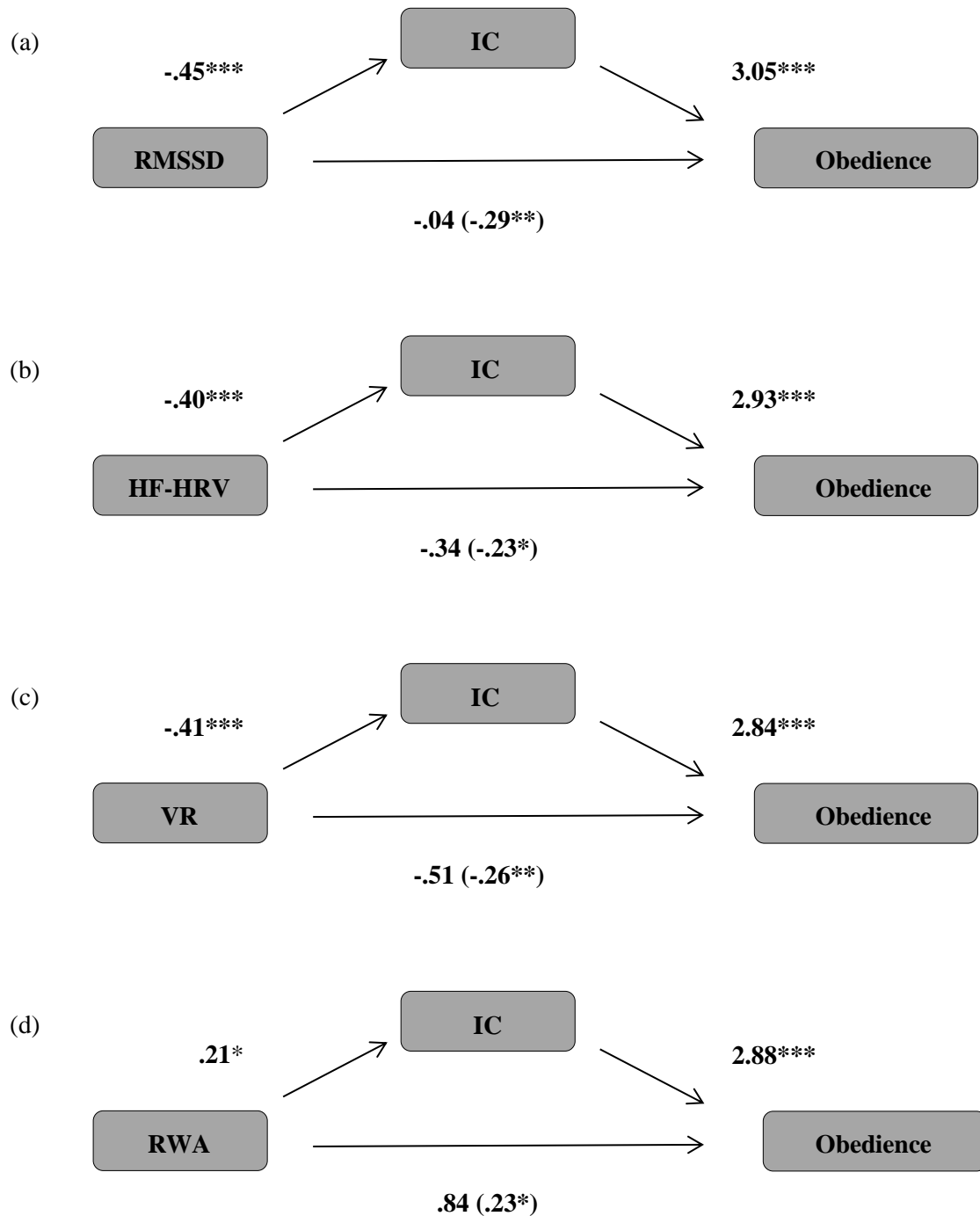


Figure 4. Standardized regression coefficients (a) for the relationship between RMSSD (at rest) and relative obedience as mediated by inhibitory control, (b) for the relationship between HF-HRV (at rest) and relative obedience as mediated by inhibitory control (c) for the relationship between VR and relative obedience as mediated by inhibitory control, (d) for the relationship between right-wing authoritarianism and relative obedience as mediated by inhibitory control. RMSSD = root mean square of successive R-R differences; HF-HRV = high frequencies (0.15-0.4 Hz) expressed in normalized units; VR = vagal reactivity; RWA = right-wing authoritarianism; IC = inhibitory control. * $p < .05$ ** $p < .01$ *** $p < .001$

Discussion

The present study first aimed to examine the influence of autonomic flexibility (i.e., higher VT at rest and higher VR during challenge) on destructive obedience. We were also interested in the emotion regulation strategies that would mediate the influence of autonomic flexibility and RWA on destructive obedience. We predicted and found that lower autonomic flexibility (i.e., lower VT at rest and lower VR) was associated with higher levels of destructive obedience. More precisely, we found that (i) lower RMSSD predicted more absolute obedience, (ii) lower RMSSD, HF-HRV and VR (indexed by decrease in HF-HRV during the first block relative to baseline) predicted higher relative obedience. We also found some evidence that obedient participants have exerted cognitive effort to inhibit their automatic emotional responses to the victim's pain (i.e., increased RMSSD and decreased heart rate relative to baseline). The reverse pattern for heart rate was found among disobedient participants (i.e., increased heart rate relative to baseline). In accordance to our expectations, inhibitory control mediated the influence of resting VT and VR on relative obedience. Finally, replicating previous findings, we found that higher RWA was associated with higher relative obedience. Results also revealed that (i) lower VT and VR predicted higher RWA scores, (ii) inhibitory control mediated the influence of RWA on relative obedience. All in all, these results suggest that autonomic rigidity and inhibition of internal responses pull for destructive obedience.

Vagal tone and destructive obedience

As mentioned previously, recent research on destructive obedience suggests that low self-regulatory control could contribute to destructive obedience (Cheetham et al., 2009; see also Zeigler-Hill, Souhard, Archer, & Donohoe, 2013). However, no studies have used ANS functioning as a biological index of self-regulation, thus leaving it unclear whether *regulation*

is “impaired” in destructive obedience per se. The association between low VT at rest and destructive obedience is consistent with the polyvagal perspective and the model of neurovisceral integration, as well as with research on empathy for pain showing that heightened stress reactivity influence one’s ability to regulate personal distress and tend to others (e.g., Eisenberg & Eggum, 2009). It is worth noting that, while the polyvagal theory and the model of neurovisceral integration both causally relate vagal activity with prosocial behavior, very few studies used behavioral measures of prosociality to establish such a relation (see Beffara, Bret, Vermeulen, & Mermillod, 2016). Our results support that vagal functioning may contribute to prosocial disobedience. According to Keltner, Kogan, Piff and Saturn (2014), the association between vagal activity and prosocial tendencies “is driven, in part, by a reduction in arousal, which enables a shift in attention to the person in need”. This is in line with research showing that attentional and emotional regulation worked together in the process of self-regulation and goal directed behaviors (e.g., Thayer & Lane, 2000). We propose that a reduced capacity to regulate the sympatho-excitatory circuits responsible for default threat responses may have limited the ability for individuals to pay attention to distress cues from the victim and to display sympathetic responses. By contrast, the ability to inhibit the default response mode to threat and to focus on a broader set of social cues may increase flexibility by decreasing stress levels and facilitate prosocial tendencies (e.g., sympathy, disobedience toward destructive orders).

Vagal reactivity and destructive obedience

The polyvagal theory highlights that “successful adaptation of mammals is dependent on the systematic and reliable withdrawal and reengagement of the vagal brake as a mechanism to rapidly regulate metabolic output in response to environmental demands” (Porges, Doussard-Roosevelt, Portales, & Greenspan, 1996). As emphasized by Muhtadie et

al. (2015), dynamic modulation of vagal influence in response to shifts in the environment would be a particularly useful trait-like physiological index that can be used to predict social sensitivity. Our study extends prior work by exploring the notion that an adaptive vagal system will exhibit greater vagal withdrawal (indexed by a decrease in HF-HRV) during challenge. The results show that individuals with greater VR are more prone to disobey destructive orders, which could suggest that vagal withdrawal is more likely to facilitate prosocial behavior under conditions that pose greater metabolic demand on individuals and that high VR individuals are more sensitive to socio-emotional information in their environment.

Spontaneous avoidance and destructive obedience

During emotion regulation, individuals may increase, maintain, or decrease positive and negative emotions. Up-regulation processes may increase the magnitude of the emotion response, down-regulation processes result in a speedier return to baseline, and maintenance processes aim to maintain the emotional response over a longer period of time (e.g., Koole, 2009). Our results are consistent with prior work showing that distressed people often regulate their emotions by resorting to avoidance, and that heightened resting VT is associated with decreased spontaneous avoidance (e.g., Aldao et al., 2015). There are also consistent (i) with pain studies showing a decreased amygdala activity that indicate a cognitive strategy to adapt to a distressful but unavoidable painful event (e.g., Petrovic et al., 2004), (ii) with moral neuroscience research suggesting that decreased amygdala responding could be responsible for decreased aversion to harming others (e.g., Blair, Mitchell, & Blair, 2005).

Right-wing authoritarianism

Finally, our results could advance our understanding of the meaning of RWA and its influence on destructive obedience. Self-regulation has long been seen as part of political

attitudes, with research suggesting that RWA is related to cognitive rigidity, threat processing and defensiveness (e.g, Duckitt, 2001). The polyvagal theory and the model of neurovisceral integration both conceptualize VT as a useful index of threat perceptions or appraisals and therefore of stress. Thus, a rigid attentional bias toward threat-related stimuli associated with a reduced capacity to alter psychophysiological responses in synchrony with situational demands may underlie the psychological motivations to endorse authoritarian beliefs, and cause difficulties to experience sympathy in response to the others' distress. Our results also highlight that high RWA individuals are more prone to inhibit their internal experiences in response to pain-related cues from the other, which is consistent with previous research showing that emotion regulation strategy of avoidance is a common maladaptive response to aversive stimuli among distressed individuals (e.g., Aldao et al., 2015).

Conclusion

A great deal of research has been devoted to explaining the causes of destructive obedience, but biological theories are rarely used to generate hypotheses and only very few studies considered biological differences as an important explanatory and predictive factor. A widely-held assumption is that destructive behaviors are entirely the product of environmental forces (see Blass, 1991) but, as emphasized by Hatemi and McDermott (2011), “who we are and what we do is the product of environmental *and* biological factors”. The presents study highlights that biology “can help to account both for central behavioral tendencies and also for the astounding variation around these central tendencies” (Hatemi & McDermott, 2011), providing empirical evidence that autonomic rigidity favors destructive obedience through the use of emotion regulation strategy of avoidance (i.e., inhibition of automatic responses to the victim's pain). In our view, the use of current models of self-regulation such as the polyvagal theory and the model of neurovisceral integration may represent a significant step forward in

understanding the mechanisms underlying destructive obedience. Following these perspectives, our results suggest that individuals with greater autonomic flexibility have better abilities to orient their attention toward pain cues from the others without automatically falling into a default mode response to threat, and thus to display sympathetic responses. Further studies might determine if attentional skills mediate the association between autonomic flexibility and destructive obedience. Moreover, we operationalized avoidance in term of inhibitory control over automatic emotional responses, so it will be interesting for future studies to investigate the role of additional forms of avoidance, such as expressive suppression. More important, it will be essential to test causal associations between autonomic flexibility and destructive obedience. Several techniques exist for experimentally inducing changes in VT, as is the case with pharmacological blockade and biofeedback training (e.g., Lehrer & Gevirtz, 2014). For instance, Bornemann, Kok, Böckler and Singer (2016) investigated the influence of deliberate regulation of HF-HRV on altruistic behavior using a biofeedback task and found that variation in the ability to voluntary upregulate HF-HRV predicts individual differences in altruistic prosocial behavior.

References

- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Akinola, M, & Mendes, W. B. (2014). It's good to be the king: Neurobiological benefits of higher social standing. *Social Psychological and Personality Science*, 5(1), 43-51. doi:10.1177/1948550613485604.
- Aldao, A., Dixon-Gordon, K. L., & De Los Reyes, A. (2015). Individual differences in physiological flexibility predict spontaneous avoidance. *Cognition and Emotion*, 30(5), 985-998. doi:10.1080/02699931.2015.1042837
- Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Canada: University of Manitoba Press.
- Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.
- Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, 10(3), 229-240. doi:10.1037/1089-2680.10.3.229
- Beauvois, J. L., Courbet, D., & Oberlé, D. (2012). The prescriptive power of the television host: A transposition of Milgram's obedience paradigm to the context of TV game show. *European Review of Applied Psychology*, 62(4), 111-119. doi:10.1016/j.erap.2012.02.001
- Beffara, B., Bret, A.G., Vermeulen, N., & Mermillod, M. (2016). Resting High Frequency Heart Rate Variability Selectively Predicts Cooperative Behavior. *Physiology & Behavior*, 164, 417-428. doi:10.1016/j.physbeh.2016.06.011

- Bègue, L., Beauvois, J. L., Courbet, D., Oberlé, D., Lepage, J., & Duke, A. (2015). Personality Predicts Obedience in a Milgram Paradigm. *Journal of Personality*, 83(3), 299-306. doi:10.1111/jopy.12104
- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, 11(9), 387-392. doi:10.1016/j.tics.2007.07.003
- Blair, R. J. R., Mitchell, D., & Blair, K. (2005). *The Psychopath: Emotion and the Brain*. Oxford, UK: Blackwell.
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychology Inquiry*, 13(2), 103-124. doi:10.1207/S15327965PLI1302_01
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, 60(3), 398-413. Retrieved from <http://www.stanleymilgram.com/pdf/understanding20%behaviour.pdf>
- Bornemann, B., Kok, B. E., Böckler, A., & Singer, T. (2016). Helping from the heart: Voluntary up-regulation of heart rate variability predicts altruistic behavior. *Biological Psychology*, 119, 54-63. doi:10.1016/j.biopsycho.2016.07.004
- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of Psychophysiology*. Cambridge, UK: Cambridge University Press.
- Cheetham, M., Pedroni, A. F., Antley, A., Slater, M., & Jäncke, L. (2009). Virtual Milgram: Empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Frontiers in Human Neuroscience*, 3(29). doi:10.3389/neuro.09.029.2009

- Dambrun, M., & Vatiné, E. (2010). Reopening the study of extreme social behaviors: Obedience to authority within an immersive video environment. *European Journal of Social Psychology, 40*(5), 760-773. doi:10.1002/ejsp.646
- Decety, J., & Lamm, C. (2009). Empathy versus Personal Distress: Recent Evidence from Social Neuroscience. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 199-213). Cambridge, MA: MIT Press.
- Decety, J., & Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience, 2*(1), 1-24. doi:10.1016/j.dcn.2011.05.003
- de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Science, 10*(10), 435-441. doi:10.1016/j.tics.2006.08.008
- Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 33, pp. 41-113). New York: Academic Press.
- Duckitt, J. (2013). Introduction to the special section on authoritarianism in societal context: The role of threat. *International Journal of Psychology, 48*(1), 1-5. doi:10.1080/00207594.2012.738298
- Eisenberg, N., & Eggum, N. D. (2009). Empathic Responding: Sympathy and Personal Distress. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71-83). Cambridge, MA: MIT Press.
- Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg & W. Damon (Eds.), *Handbook of child psychology. Vol. 3. Social, emotional, and personality development*. 6th edition (pp. 646-718). New York: Wiley.
- Eisenberg, N., Valiente, C., & Champion, C. (2004). Empathy-Related Responding: Moral, Social, and Socialization Correlates. In A. G. Miller (Ed.), *The Social Psychology of*

- Good and Evil: Understanding Our Capacity for Kindness and Cruelty* (pp. 386-415).
New York: Guilford Press.
- Elms, A. C. (1995). Obedience in Retrospect. *Journal of Social Issues*, 51(3), 21-31.
doi:10.1111/j.1540-4560.1995.tb01332.x
- Elms, A. C., & Milgram, S. (1966). Personality characteristics associated with obedience and defiance toward authoritative command. *Journal of Experimental Research in Personality*, 1(4), 282-289.
- Fabes, R. A., & Eisenberg, N. (1997). Regulatory control and adults' stress-related responses to daily life events. *Journal of Personality and Social Psychology*, 73(5), 1107-1117.
Retrieved from
http://www.researchgate.net/profile/Nancy_Eisenberg/publication/13864629_Regulatory_control_and_adults%27_stress_related_responses_to_daily_life_events/links/02e7e539be941aade6000000.pdf
- Fabes, R. A., Eisenberg, N., & Eisenbud, L. (1993). Behavioral and physiological correlates of children's reactions to others' distress. *Developmental Psychology*, 29(4), 655-663.
doi:10.1037/0012-1649.29.4.655
- Feldman, S., Lavine, H., Lodge, M., & Verhulst, B. (2010). *Seeing Negative: Authoritarianism and Automatic Vigilance for Threatening Stimuli*. Paper presented at the Annual Meeting of the International Society of Political Psychology, San Francisco, CA.
- Goubert, L., Craig, K. D., & Buysse, A. (2009). Perceiving Others in Pain: Experimental and Clinical Evidence on the Role of Empathy. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 153-166). Cambridge, MA: MIT Press.
- Gross, J. J. (2002). Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology*, 39(3), 281-91. doi:10.1017.S0048577201393198

- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implication for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348-362. doi:10.1037/0022-3514.85.2.348
- Hatemi, P. K., & McDermott, R. (2011). Evolution as a Theory for Political Behavior. In P. K. Hatemi & R. McDermott (Eds.), *Man is by Nature a Political Animal: Evolution, Biology, and Politics*. (pp. 13-46). Chicago, IL: University of Chicago Press.
- Heathers, J. A. (2014). Everything Hertz: methodological issues in short-term frequency-domain HRV. *Frontiers in Physiology, 5*, 177. doi:10.3389/fphys.2014.00177
- Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences, 37*(3), 297-350. doi:10.1017/S0140525X13001192
- Iorfino, F., Alvares, G. A., Guastella, A. J., & Quintana, D. S. (2015). Cold face test-induced increases in heart rate variability are abolished by engagement in a social cognition task. *Journal of Psychophysiology, 30*, 38-46. doi:10.1027/0269-8803/a000152
- Jost, J. T., & Amodio, D. M. (2012). Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion, 36*(1), 55-64. doi:10.1007/s11031-011-9260-7
- Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of Obedience: Toward a Social Psychology of Authority and Responsibility*. New Haven: Yale University Press.
- Keltner, D., Kogan, A., Piff, P. K., & Saturn, S. R. (2014). The sociocultural appraisals, values, and emotions (SAVE) framework of prosociality: core processes from gene to meme. *Annual Review of Psychology, 65*, 425-460. doi:10.1146/annurev-psych-010213-115054
- Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion, 23*(1), 4-41. doi:10.1080/02699930802619031

- Lankford, A. (2009). *Human Killing Machines: Systematic Indoctrination in Iran, Nazi Germany, Al Qaeda, and Abu Ghraib*. Lanham, MD: Lexington Books.
- Lehrer, P. M., & Gevirtz R. (2014). Heart rate variability biofeedback: how and why does it work? *Frontiers in Psychology, 5*, 756. doi:10.3389/fpsyg.2014.00756
- Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371-378. doi:10.1037/h0040525
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. London: Tavistock Publications.
- Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science*. New York: Praeger.
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Science, 11*(8), 319-321. doi:10.1016/j.tics.2007.06.001
- Montano, N., Porta, A., & Malliani, M. (2001). Evidence for Central Organization of Cardiovascular Rhythms. *Annals of the New York Academy of Sciences, 940*(1), 299-306. doi:10.1111/j.1749-6632.2001.tb03685.x
- Movius, H. L., & Allen, J. J. B. (2005). Cardiac vagal tone, defensiveness, and motivational style. *Biological Psychology, 68*(2), 147-162. doi:10.1016/j.biopsycho.2004.03.019
- Muhtadie, L., Koslov, K., Akinola, M., & Mendes, W. B. (2015). Vagal flexibility: A physiological predictor of social sensitivity. *Journal of Personality and Social Psychology, 109*(1), 106-120. doi:10.1037/pspp0000016
- Negrao, B. L., Bipath, P., Van der Westhuizen, D., & Viljoen, M. (2011). Autonomic correlates at rest and during evoked attention in children with attention-deficit/hyperactivity disorder and effects of methylphenidate. *Neuropsychobiology, 63*(2), 82-91. doi:10.1159/000317548

- Obradović, J., Bush, N. R., Stamperdahl, J., Adler, N. E., & Boyce, W. T. (2010). Biological sensitivity to context: The interactive effects of stress reactivity and family adversity on socioemotional behavior and school readiness. *Child Development, 81*(1), 270-289. doi:10.1111/j.1467-8624.2009.01394.x.
- Otzenberger, H., Gronfier, C., Simon, C., Charloux, A., Ehrhart, J., Piquard, F., et al. (1998). Dynamic heart rate variability: a tool for exploring sympathovagal balance continuously during sleep in men. *The American Journal of Physiology, 275*(3 pt 2), H946-H950. Retrieved from <http://ajpheart.physiology.org/content/275/3/H946.long>
- Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., et al. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research, 59*(2), 178-193. doi:10.1161/01.RES.59.2.178
- Park, G., & Thayer, J. F. (2014). From the heart to the mind: cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Frontiers in Psychology, 5*, 278. doi:10.3389/fpsyg.2014.00278
- Park, G., Vasey, M., Van Bavel, J. J., & Thayer, J. F. (2013). Cardiac vagal tone is correlated with selective attention to neutral distractors under load. *Psychophysiology, 50*(4), 398-406. doi:10.1111/psyp.12029
- Petrovic, P., Carlsson, K., Petersson, K. M., Hansson, P., & Ingvar, M. (2004). Context-dependent deactivation of the amygdala during pain. *Journal of Cognitive Neuroscience, 16*(7), 1289-1301. doi:10.1162/0898929041920469
- Porges, S. W. (1992). Vagal tone: A physiological marker of stress vulnerability. *Pediatrics, 90*(3), 498-504. Retrieved from http://www.researchgate.net/publication/21641337_Vagal_tone_a_physiologic_marker_of_stress_vulnerability

- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage: A Polyvagal Theory. *Psychophysiology*, *32*(4), 301-318. doi:10.1111/j.1469-8986.1995.tb01213.x
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, *74*(2), 116-143. doi:10.1016/j.biopsycho.2006.06.009
- Porges, S. W., Doussard-Roosevelt, J. A., Portales, A. L., & Greenspan, S. I. (1996). Infant regulation of the vagal “brake” predicts child behavior problems: A psychobiological model of social behavior. *Developmental Psychobiology*, *29*(8), 697-712. doi:10.1002/(SICI)1098-2302(199612)29:8<697::AID-DEV5>3.0.CO;2-O
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717-731. doi:10.3758/BF03206553
- Quintana, D. S., & Heathers, J. A. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology*, *5*, 805. doi:10.3389/fpsyg.2014.00805
- Segerstrom, S. C., & Ness, S. L. (2007). Heart rate variability indexes self-regulatory strength, effort, and fatigue. *Psychological Science*, *18*(3), 275-281. doi:10.1111/j.1467-9280.2007.01888.x
- Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, *1156*, 81-96. doi:10.1111/j.1749-6632.2009.04418.x
- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *Plos One*, *1*(1), e39. doi:10.1371/journal.pone.0000039

- Stellar, J. E., Cohen, A., Oveis, C., & Keltner, D. (2015). Affective and physiological responses to the suffering of others: Compassion and vagal activity. *Journal of Personality and Social Psychology, 108*(4), 572-585. doi:10.1037/pspi0000010
- Thayer, J. F., Ahs F., Fredrikson, M., Sollers J. J., III, & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews, 36*(2), 747-756. doi:10.1016/j.neubiorev.2011.11.009
- Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine, 37*(2), 141-153. doi:10.1007/s12160-009-9101-z
- Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders, 61*(3), 201-216. doi:10.1016/S0165-0327(00)00338-4
- Thayer, J. F., & Lane, R. D. (2007). The role of vagal function in the risk for cardiovascular disease and mortality. *Biological Psychology, 74*(2), 224-242. doi:10.1016/j.biopsycho.2005.11.013
- Van Roon, A. M., Mulder, L. J. M., Althaus, M., & Mulder, G. (2004). Introducing a baroreflex model for studying cardiovascular effects of mental workload. *Psychophysiology, 41*(6), 961-981. doi:10.1111/j.1469-8986.2004.00251.x
- Waller, J. (2002). *Becoming Evil: How Ordinary People Commit Genocide and Mass Killing*. Oxford: Oxford University Press.
- Weissflog, M. J., van Noordt, S. J. R., Choma, B. L., Dywan, J., & Segalowitz, S. J. (2010). Sociopolitical ideology and electrocortical responses. *Psychophysiology, 47*(Sup), S24.

Zeigler-Hill, V., Southard, A. C., Archer, L. M., & Donohoe, P. L. (2013). Neuroticism and Negative Affect Influence the Reluctance to Engage in Destructive Obedience in the Milgram Paradigm. *Journal of Social Psychology, 153*(2), 161-174.
doi:10.1080/00224545.2012.713041

Flexibilité physiologique et obéissance destructrice

Les résultats de l'étude 3 montrent (i) que l'OD est associée à une moindre flexibilité physiologique, (ii) que les participants obéissants ont exercé un effort cognitif couplé à une diminution du stress physiologique, (iii) que cette diminution du stress physiologique médiatise les effets de la flexibilité physiologique et du RWA sur l'OD. Ces résultats apportent un premier soutien empirique à l'hypothèse voulant que la vulnérabilité au stress influence l'OD *via* le contrôle inhibiteur de la résonance empathique. L'étude 4 avait pour objectif de tester plus directement l'hypothèse d'une relation de causalité entre contrôle inhibiteur et OD.

4.2. Obéissance destructrice et « fatigue mentale »

Objectifs et présentation de l'étude

Les résultats d'un grand nombre de travaux montrent qu'il est possible d'induire expérimentalement une « fatigue mentale » responsable d'un affaiblissement temporaire du contrôle inhibiteur. Par exemple, des études ont montré que la réalisation d'une première tâche requérant un contrôle cognitif important diminue le contrôle inhibiteur à une tâche subséquente (e.g., Dang, 2017). Baumeister et ses collaborateurs ont proposé que l'autocontrôle soit une ressource limitée, et que la diminution des ressources d'autocontrôle affecte négativement la décision morale en augmentant l'influence des processus émotionnels automatiques (e.g., Baumeister, Sparks, Stillman, & Vohs, 2008). De récents travaux ont remis en cause cette approche (voir Milyavskaya & Inzlicht, 2016), et nous avons vu dans l'introduction du présent travail doctoral le rôle important des processus émotionnels automatiques dans le raisonnement moral orienté vers le bien-être d'autrui (e.g., Blair, 2007 ; Crockett et al., 2010 ; Decety & Lamm, 2009). Inzlicht et ses collaborateurs ont proposé que (i) la réalisation d'une première tâche requérant un contrôle cognitif important diminue la motivation à exercer un effort similaire à une tâche subséquente, (ii) les réponses automatiques ne sont pas intrinsèquement « bonnes » ou « mauvaises » (e.g., Inzlicht, Schmeichel, & Macrae, 2014). Selon l'approche développée par les auteurs, (i) un « dilemme d'autocontrôle » met en tension une réponse automatique et une réponse contrôlée visant la satisfaction d'un but distant, (ii) une baisse de motivation à l'autocontrôle augmente la valeur subjective des réponses automatiques, (iii) la motivation oriente l'attention et affecte la perception. Dans le paradigme d'obéissance, une diminution de la motivation à l'autocontrôle pourrait orienter l'attention des participants vers la détresse de la victime et favoriser une meilleure compréhension de son expérience interne, augmentant alors la valeur subjective des

réponses automatiques (e.g., aversion pour l'atteinte à autrui). Ainsi, la fatigue mentale pourrait diminuer l'OD. Il est également raisonnable de faire l'hypothèse d'un effet d'interaction entre fatigue mentale et RWA. Le contrôle inhibiteur des réponses émotionnelles automatiques pourraient expliquer la plus forte propension à l'OD des personnes présentant de forts scores de RWA. Cette stratégie d'autorégulation basée sur l'évitement pourrait refléter un effort cognitif visant à réduire la détresse personnelle. La valorisation de l'autorité pourrait également contribuer à la motivation au contrôle inhibiteur. Selon Milyavskaya et Inzlicht (2016), lorsqu'un « dilemme d'autocontrôle » met en tension une réponse automatique et une réponse contrôlée de valeur égale, il est nécessaire d'exercer un effort à l'autocontrôle pour favoriser le but distant. Les personnes présentant de forts scores de RWA pour être plus disposées à faire l'expérience d'un conflit entre réponse automatique et réponse contrôlée, alors que les personnes présentant de faibles scores de RWA pour être plus enclines à orienter leur attention vers la détresse de la victime et à favoriser les réponses automatiques. Cette dernière proposition est cohérente avec les travaux en psychologie de la morale suggérant (i) que la moralité humaine s'articule autour de cinq dimensions (« care », réciprocité, loyauté envers le groupe d'appartenance, respect de l'autorité, pureté morale), (ii) que les intuitions morales des personnes progressistes reposent en priorité sur le « care » et l'aversion pour l'atteinte à autrui (e.g., Haidt & Graham, 2007).

Avant leur participation à l'étude, les participants ($n = 99$) ont répondu à un questionnaire en ligne comprenant une version française de l'échelle RWA. Préalablement à la procédure d'obéissance, les participants ont été aléatoirement répartis dans deux conditions expérimentales (« Faible effort » vs « Effort élevé »). Dans la condition « Faible effort », les participants ont réalisé une version peu exigeante de la tâche de « détection de lettres » (e.g., Baumeister, Bratslavsky, Muraven, & Tice, 1998), alors que les participants de la condition « Effort élevé » ont eu à réaliser une version plus coûteuse en termes d'effort cognitif.

Manuscrit 4⁹

⁹ Manuscrit accepté par la revue *Cognition and Emotion*

Influence of Authoritarianism, Vagal Tone and Mental Fatigue on Obedience to Authority

Johan Lepage¹, Laurent Bègue¹, Oulmann Zerhouni², Rémi Courset¹ & Martial Mermillod^{3,4}

¹ Laboratoire InterUniversitaire de Psychologie, Département de psychologie, Université Grenoble Alpes, Grenoble, France

² Laboratoire Parisien de Psychologie Sociale, Département de psychologie, Université Paris Ouest Nanterre La Défense, Nanterre, France

³ Laboratoire de Psychologie et Neurocognition, CNRS UMR 5105, Département de psychologie, Université Grenoble Alpes, Grenoble, France

⁴ Institut Universitaire de France, Paris, France

Corresponding author

Johan Lepage, PhD Candidate: Laboratoire InterUniversitaire de Psychologie, UFR SHS, 1251 avenue Centrale, BP 47, 38040 Grenoble Cedex 9, France; E-mail: johan.lepage@gmail.com, Phone: (+33 4) 76 82 58 92, Fax: (+33 4) 76 82 56 65

Abstract

Recent research suggests that obedience in the Milgram paradigm is underpinned by stress vulnerability and inhibitory control over pain sharing. Because self-regulatory fatigue (SRF) induction is a suited method to investigate the influence of inhibitory control on behavior, participants ($n = 99$) were randomly assigned to a High vs. Low self-regulatory condition. Heart rate variability (HRV, a biomarker of stress vulnerability) was collected during 5-min baseline and continuously during the experimental procedure. Prior to the experiment, participants completed an online survey assessing right-wing authoritarianism (RWA), a well-known predictor of obedience. Using the Immersive Video Milgram Obedience Experiment, we found (i) that lower resting HRV predicted higher destructive obedience, (ii) that low self-regulatory inhibition (induced by fatigue) reduced destructive obedience, (iii) that the well-established influence of RWA on destructive obedience was suppressed in the presence of SRF. Implications for future directions in obedience research are discussed.

Keywords

Obedience to authority, right-wing authoritarianism, mental fatigue, vagal tone, inhibitory control

Introduction

It is now more than 50 years since Milgram published his first report of his experimental study of obedience to authority (Milgram, 1963). Milgram's obedience experiments (Milgram, 1963, 1965, 1974) have achieved a remarkable visibility and are still widely cited when explaining destructive obedience (i.e., obedience to orders that cause harm to another individual; e.g., Kelman & Hamilton, 1989; Lankford, 2009; Mermillod, Marchand, Lepage, Bègue, & Dambrun, 2015). These experiments also ignited an intense ethical controversy and the evolution of ethical standards has discouraged researchers from extending Milgram's work (e.g., Elms, 1995). As a result the mechanisms responsible for destructive obedience remain unknown (e.g., Blass, 1991; Cheetham, Pedroni, Antley, Slater, & Jäncke, 2009). However, experimental research was recently reopened through the use of immersive environments and new data point out the role of self-regulation processes.

Milgram's obedience experiments

The Milgram paradigm consists of ordering a naïve participant to administer electric shocks to a victim (a confederate) in the context of a learning experiment ostensibly set up to study the effects of punishment on memory. A simulated shock generator is used with 30 graded switches ranging from 15 to 450 volts. The generator bears verbal designation ranging from "Slight Shock" to "Danger: Severe Shock". As the experiment proceeds, the participant is ordered to administer shocks of increasing voltages to the victim whenever he gave a wrong answer in a word-memory test. A pool of Yale seniors (all psychology majors) predicted that only an extreme minority (approximately 1.2%) would continue through to the most severe shock (450 volts). Contrary to expectations, 65% of participants continued to give "shocks" to the maximum 450 volts, in spite of the victim's protests (condition 5).

Current theories of destructive obedience emphasize on cognition and rationality and downplay emotions as guides of human behavior. For instance, both the agentic state model (Milgram, 1974) and the socio-cognitive theory (e.g., Bandura, 1999) emphasized on the disengagement of cognitive mechanisms responsible for moral control. In contrast, social neuroscience research highlights that care-based moral reasoning and prosocial behavior (i.e., actions that are intended to help or benefit another individual; Decety & Svetlova, 2012) rely on subcortical processes of pain-related affective sharing and emotion regulation (e.g., Decety & Lamm, 2009; Eisenberg & Eggum, 2009;). A recent partial replication of Milgram's paradigm within an immersive virtual environment threw some light on the neural and affective underpinnings of destructive obedience.

Destructive obedience and personal distress

Milgram (1974) noted that almost all participants exhibited signs of emotional distress, but whether the participants' response to the victim's pain was more akin to other-oriented empathic concern or to a self-oriented aversive state of personal distress is unclear (Cheetham et al., 2009). Research on empathy for pain demonstrates that pain-related affective sharing can provoke either sympathy (an other-oriented motivated response) or personal distress (an aversive, self-oriented motivated response; e.g., Decety & Lamm, 2009). While sympathy initiates care-based moral reasoning and prosocial behavior, personal distress is associated with the desire to alleviate one's own but not the other's distress (Eisenberg & Eggum, 2009). A large range of studies showed that individuals with low self-regulation abilities are biased to experience over-arousal and personal distress in response to the others' pain (e.g., Eisenberg et al., 1994, 2006). Studies also found that empathy-related responses are malleable with respect to several factors such as contextual appraisal (e.g., de Vignemont & Singer, 2006).

The Milgram paradigm induces a social dilemma in which participants are wedged between conflicting alternatives “for which there is no prospect of establishing a personally and collectively favorable outcome” (Cheetham et al., 2009). Using a virtual version of the Milgram paradigm, the authors found that the observation of the victim’s pain within this context induced brain activations related to an aversive, self-oriented state of personal distress (i.e., activations in pre-defined brain areas involved in affective processing - e.g., right amygdala - but not in those commonly associated with emotion regulation - e.g., anterior cingulate cortex). These findings suggest that low self-regulatory control of the shared affect evoked by the emotional state of the victim could be responsible (at least partially) for destructive obedience.

Emotion down-regulation processes

There is a growing appreciation that distressed individuals use a wide range of strategies for down-regulating aversive emotions (Gross, 2002; Gross & John, 2003). Emotion regulation strategy of avoidance (e.g., inhibition of internal responses, cognitive reappraisal) is a common maladaptive response to aversive emotional experiences (e.g., Aldao, Dixon-Gordon, & De Los Reyes, 2015). According to Goubert, Craig and Buysse (2009), individuals responding with high distress to the other’s pain could be motivated “to underestimate the observed person’s pain, in an attempt to keep their own distress within acceptable limits”. As a consequence, individuals may tend not to fully attend to the other’s experience and fail to display sympathetic responses. Brain imaging studies found a decreased amygdala activity during the observation and the experience of pain that indicate a cognitive strategy to attenuate pain-related stress responses in a “no-escape situation” (e.g., Petrovic, Carlsson, Petersson, Hansson, & Ingvar, 2004). Because decreased activity within the amygdala has been linked to decreased harm aversion (e.g., Blair, 2007), such an inhibitory control could

have detrimental consequences in a Milgram-like paradigm (i.e., increased destructive obedience).

All in all, previous research suggests that inhibitory control over pain sharing pulls for destructive obedience while automatic emotional processes pull for disobedience toward destructive orders. The direct experimental manipulation of the individuals' capacity for effortful self-regulation may represent a significant step forward in understanding the role of self-regulation in the Milgram paradigm. More precisely, inducing self-regulatory fatigue (SRF) should decrease destructive obedience.

Self-regulatory fatigue

A plethora of studies using a sequential task paradigm demonstrate that exerting high self-regulatory effort (SRE) at Time 1 results in self-regulatory impairment at Time 2 (see Dang, 2017, for an updated meta-analysis). The predominant explanation of such effect claims that self-control is a limited resource that would get depleted with use (e.g., Muraven & Baumeister, 2000) but recent findings are incompatible with the resource model. For instance, studies found that allowing a rewarding experience in the interval between two self-control tasks prevents reduced self-control in the second task (see Milyavskaya & Inzlicht, 2016). A widely-held assumption associated with the resource model of self-control is that SRF would negatively affect decision making by enhancing the role of automatic emotional processes over cognitive processes (e.g., Baumeister, Sparks, Stillman, & Vohs, 2008). But a growing body of data highlights that automatic aversive reactions to harms infuse moral judgment and motivate prosocial behavior (e.g., Blair, 2007; Crockett, Clark, Hauser, & Robbins, 2010; Damasio, Tranel, & Damasio, 1990). Inzlicht and colleagues recently challenged the resource model with a non-resource based process model. The “shifting priorities model” of self-control (Inzlicht, Schmeichel, & Macrae, 2014) relies on a

motivation-shift mechanism: after exerting mental effort, the value of exerting further effort becomes reduced, while the value of giving in to the automatic impulse is increased, such that the latter predominates (Kool & Botvinick, 2014). The shifting priorities model also emphasizes that automatic impulse is not inherently “good” or “bad”. This model posits that (i) self-control dilemmas involve an immediate impulse (e.g., to avoid causing harm) that competes with a distal goal (e.g., to produce beneficial knowledge to society), (ii) shifts in motivation are responsible for changes in the subjective values of each option, (iii) motivation drives attention and can affect perception, thereby influencing the valuation process (i.e., the process of integrating inputs related to the subjective values of each option; Milyavskaya & Inzlicht, 2016).

If destructive obedience within the Milgram paradigm requires substantial cognitive effort to inhibit empathic resonance, fatigued individuals may be unwilling to do so and chose to forgo further inhibitory control in favor of automatic emotional responses to the victim’s pain and distress. In other words, low self-regulatory inhibition (induced by fatigue) should reduce destructive obedience, because effortful self-regulation would be needed to disengage from the pain observed on the recipient of shocks.

Inter-individual differences

Most reports of Milgram’s experiments emphasized on situational determinants of destructive obedience and downplayed the importance of individual differences (e.g., Bègue et al., 2015). That there are important inter-individual differences in destructive obedience is a fact (see Blass, 1991) and few studies indicate that authoritarianism can predict destructive obedience (e.g., Altemeyer, 1981; Dambrun & Vatiné, 2010; Elms & Milgram, 1966).

Right-wing authoritarianism

Authoritarianism seems clearly theoretically relevant to destructive obedience. Adorno, Frenkel-Brunswik, Levinson and Sanford (1950) first found that generalized prejudice, ethnocentric in-group valorisation, politico economic conservatism and pro-fascist attitudes covaried to form a “general attitudinal syndrome” termed “authoritarian personality”. According to the authors, this personality dimension consists in nine covarying traits which they attempted to measure with their F-scale (F for fascist). The Altemeyer’s right-wing authoritarianism (RWA) scale (Altemeyer, 1981, 1988) incorporated a reconceptualization of authoritarianism as well as important psychometric refinements. The items capture three facets of authoritarianism that the author found to cohere: (i) authoritarian submission (a submissive, uncritical attitudes toward authorities), (ii) authoritarian aggression (tendency to condemn, reject, and punish people who violate conventional values), and (iii) conventionalism (rigid adherence to conventional middle-class values). Elms and Milgram (1966) first found that fully obedient participants scored higher on the F-scale (Adorno et al., 1950). Other studies found that participants scoring higher on the RWA scale were those who administered more electric shocks (e.g., Altemeyer, 1981; Dambrun & Vatiné, 2010). To the best of our knowledge, previous studies did not investigate the mechanisms responsible for the influence of RWA on destructive obedience.

In line with our idea that destructive obedience requires substantial effortful control, one could hypothesize that high-RWA individuals are more prone to comply with destructive orders because they are more willing to use avoidance in response to the others’ pain. Seeing political attitudes through the lens of self-regulation has been spurred by research linking threat processing with ideology. Authoritarianism has been first theorized to involve threat sensitivity and cognitive rigidity. For instance, Allport (1954) argued that authoritarianism is characterized by insecurity and fearfulness. These feelings would lead the authoritarian personalities (i) to incur a need structure, order and control in their social environments, and

(ii) to react with punitive hostility to unconventionality, novelty and change. Experimental and longitudinal research demonstrates the influence of social threat (e.g., economic crises, insecurity): (i) a perceived threat increases RWA at both individual and societal levels, (ii) high-RWA individuals are more reactive to threat, and (iii) experimental manipulation of threat influence RWA scores (see Duckitt, 2013). According to Duckitt and colleagues (e.g., Duckitt, 2013; Duckitt & Sibley, 2010), high RWA expresses the motivational goal and value of establishing or maintaining collective security, which are made chronically salient by the belief that the social world is inherently dangerous and threatening. With regard to implicit tendencies, RWA has been associated with automatic recognition of negative words suggesting threats to social order and group cohesion (Feldman, Lavine, Lodge & Verhulst, 2010). Thus, RWA may reflect individual differences in threat sensitivity. In other words, chronic difficulties in inhibiting the default response mode to threat (e.g., Brosschot, Verkuil, & Thayer, 2016), in down-regulating subcortical threat circuits, would cause individuals to endorse right-wing authoritarian beliefs. Brain imaging studies demonstrate (i) that the perception of the others' pain results, in the observer, in the activation of the neural network involved in the processing of firsthand experience of pain, (ii) that the shared neural representations in the pain matrix are not specific to the experience of pain, but instead are associated with more general mechanisms such as aversion and withdrawal (see Decety & Lamm, 2009). The discovery that the observation of the others' pain activates subcortical circuits involved in threat processing (e.g., amygdala) has important implications for the question of whether observing another's distress will result in other- versus self-oriented responses. Because high-RWA individuals seem to be more sensitive to threat-related stimuli, they could be biased to experience personal distress and to use avoidance in response to the others' pain. Within the Milgram paradigm, SRE among high-RWA individuals could be a cognitive strategy to attenuate stress responses.

Alternatively, high-RWA individuals could be more likely to engage effortful self-control because they value authority. Milyavskaya and Inzlicht (2016) posit that in case where a decision relates to a conflict between similarly valued automatic impulse and distal goal, effortful self-control needs to be exercised to favor the distal goal. High-RWA individuals could be more prone to experience conflict, while low-RWA individuals could be more willing to pay attention to pain cues from the victim and to use less control to refrain automatic responses. This is consistent with research suggesting that there are five psychological foundations of morality (harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, purity/sanctity) and that liberals have moral intuitions primarily based upon caring about and reacting emotionally to harm (e.g., Haidt & Graham, 2007).

Stress vulnerability

As mentioned above, Cheetham et al. (2009) found that pain sharing within the context of the obedience dilemma elicited a self-oriented aversive state of personal distress. Could inter-individual differences in stress vulnerability explain inter-individual differences in destructive obedience? Individuals differ in their vulnerability to stressors but the mechanisms responsible for this variability are not well understood. Studies point to baseline psychophysiological characteristics, such as resting heart rate variability (HRV), as important variables linked to the individual capacity to face a stressful situation. HRV is a widely used index of cardiac vagal tone and biomarker of autonomic flexibility and adaptability (e.g., Porges, 2007). A flexible autonomic nervous system allows for rapid generation or modulation of physiological and emotional states in synchrony with changes in the environment (e.g., Appelhans & Luecken, 2006). A host of studies have shown that higher resting HRV predicts greater regulatory control and decreased negative emotional arousal in response to stressors (Eisenberg & Eggum, 2009), while lower resting HRV is associated with

personal distress, greater social anxiety, cognitive rigidity and defensiveness (e.g., Fabes, Eisenberg, & Eisenbud, 1993; Movius & Allen, 2005).

Hypotheses

First, we expected that higher RWA scores and lower resting HRV would predict more destructive obedience. *Second*, we predicted that inducing SRF would reduce destructive obedience. As noted above, High-RWA individuals are more prone to obey in the Milgram paradigm. In terms of mechanisms, high-RWA individuals may be more motivated to invest effort in inhibitory control over pain sharing. Thus, we also predicted an interaction effect between self-regulation condition (low vs. high SRE) and RWA scores: SRF would decrease obedience among high-RWA individuals. Based on previous findings, we also predicted increased task difficulty and increased HRV after high SRE relative to low effort (e.g., Segerstrom & Ness, 2007).

Material and method

Participants

99 healthy adults (63 females, $M_{age} = 25$, $SD = 4.08$) volunteered for the experiment in exchange for 30€. Power analysis was computed with G*Power 3.1.9.2 “ANOVA: Fixed Effect, special, main effect and interactions” module. We chose a middle to high effect size for the SRF induction ($d = .62$). Power analysis showed that we needed 90 participants for a power level of .80 ($f = 0.30$, $\alpha = 0.05$, Power = .80)¹. For the effect of RWA, we based our analysis on results from Dambrun and Vatiné (2010), who found a Pearson correlation between RWA and obedience around $r = .40$. We chose a high effect size ($f = .43$) and set a high level of a priori power (Power = .90). Power analysis showed that we needed 59 participants for a power level of .90 ($f = .43$, $\alpha = .05$, Power = .90). Participants were French

or bilingual in French. They provided written informed consent before participation. The study was part of a wider research project on the role of self-regulation processes in the Milgram paradigm, reviewed and approved by the University human ethics committee from Grenoble, France (Grenoble ethics committee notice number 2013-07-04-23). The protocol was carried out in accordance with the Helsinki Declaration, Convention of the Council of Europe on Human Rights and Biomedicine.

To be eligible, participants had to be aged between 18 and 50 years, with a normal or normal-to-corrected vision, explicitly reported an absence of psychiatric, neurologic, hormonal, or cardio-vascular disease, and with no medical treatment (with the exception of contraception). Individuals who might be familiar with obedience research were not included (participants were asked if they were familiar with several famous social psychology experiments, including Milgram's studies and "Le Jeu de la Mort", a transposition of the Milgram paradigm to the context of a TV game show, broadcasted in 2010; Beauvois, Courbet, & Oberlé, 2012). 21 individuals were excluded based on this criterion. Smoking, energizing drinks (e.g., coffee, tea, etc.) and psychotropic substances (e.g. alcohol, cannabis, etc.) were prohibited to each participant the day of the experiment. They had also to avoid eating or drinking (water was allowed) during the 2 hours preceding the experiment in order to limit the influence of digestion on autonomic functioning but they had to eat in the morning (more than 2 hours before the experiment) in order to avoid fasting states (short term HRV measurement can be biased by the digestion of food since viscera are innervated by the autonomic nervous system; e.g., Quintana & Heathers, 2014). Finally, they had to avoid water consumption 30 minutes before the experiment and were asked to empty their bladder before starting the experiment.

R-to-R measurement

HRV (measured as the root mean squared successive differences in the interbeat interval; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996) was recorded during 10 minutes. Three Ag-AgCl electrodes were attached as follows: one electrode (Negative) on right clavicle and one electrode (Positive) on the left 4th or 5th rib mid-clavicular to axillary line, with the Ground on right 4th or 5th rib mid-clavicular to axillary line. ECG (sampling rate 1000 Hz) was recorded using Single BIO Amp hardware in conjunction with PowerLab 16/32 hardware and LabChart 8 software (ADInstruments Ltd, UK).

Procedure

Recruitment and questionnaire. Participants were recruited through newspaper advertisements, public notices, flyers and direct email solicitation, to participate in a research on “memory and cardiac function”. Those interested were assessed for eligibility through an online form. Then, eligible participants were asked to fill out an online survey including 10 items ($\alpha = .83$) from the French translation of Altemeyer’s 20-item (1988) RWA Scale (1 = *totally disagree* to 7 = *totally agree*).

HRV baselines and Self-regulation conditions. Participants were greeted by the experimenter (30 years-old male in a white technician’s coat). The experiment took place in a quiet and dark room; participants were seated on a comfortable chair. They were told that they will participate to two experiments: a first one about cardiac function, and a second one about memory. After being fully informed about the “first experiment”, participants gave their written consent for the participation. Participants were told that their cardiac activity will be recorded during 10 minutes and that they should try to relax. The 5 first minutes allowed the participant to shift in a quiet state and the HRV baseline was recorded during the 5 following

minutes. When this step ended, participant completed the fatiguing task (i.e., crossing-out-letters task). All the participants were given a sheet of printed text from a quantum mechanics textbook and were instructed to cross out each occurrence of the letter *e*. Subsequently, they were given another sheet of printed text similar to the first one. Participants assigned to the “low SRE” condition were told to do the same. Participants assigned to the “high SRE” condition were instructed to cross out each occurrence of the letter *e*, except if this letter is followed by a vowel or is embedded in a word in which a vowel appears two letters earlier. Participants were given a maximum of 15 minutes. Then, they were asked to evaluate the task difficulty on a 1 (not at all) to 7 (totally) scale. Finally, a second HRV baseline was recorded during 5 minutes.

Obedience paradigm. We used an alternative version of the Immersive Video Milgram Obedience Experiment (IVMOE; Dambrun & Vatiné, 2010). Recent studies have found immersive virtual environment technology useful for studying social behavior (e.g., Blascovich et al., 2002), and Slater et al. (2006) first carried out a replication of Milgram’s experiment using an immersive virtual environment. While the victim was a virtual character, the results showed that the participants responded to the situation at the behavioral, physiological and subjective levels as if it was real. The IVMOE is very similar in content to the original Milgram paradigm except that the victim (an actor filmed and recorded) is seen onscreen. Using their immersive environment, the authors replicated several important findings observed in real environments (e.g., increased anxiety, effect of RWA on obedience, displacement of responsibility among obedient participants).

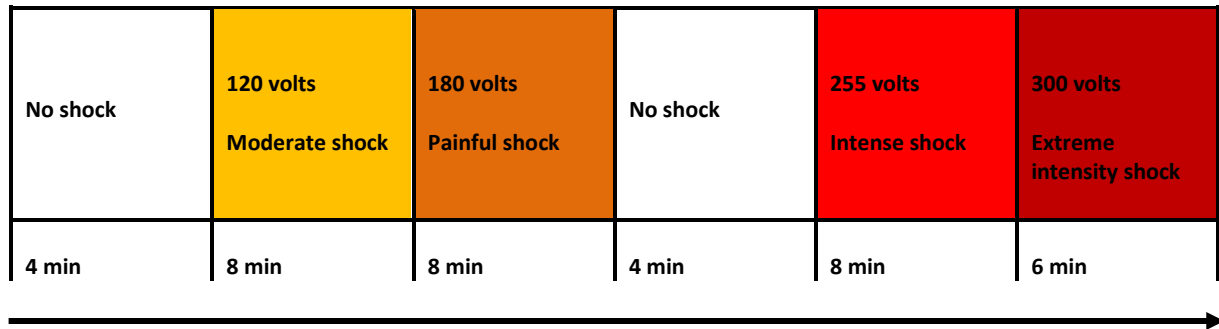
First, participants in our study were told that the experiment was a simulation, as follow: “This experiment will be role-play. You will interact with another person; this is an actor that will simulate reactions to the experimental procedure. We ask you to behave exactly

as if everything was real. You will have to make decisions during the experiment; we ask you to make these decisions not by considering that it will have no consequences because of the role-play, but by considering that each of your decision will have concrete consequences on real persons within a real situation. This is the most important point: you have to behave as if everything was real. If you are ok with this, right now I will behave myself as if everything was real. If you ask me about the role play, I will respond that everything is real.”

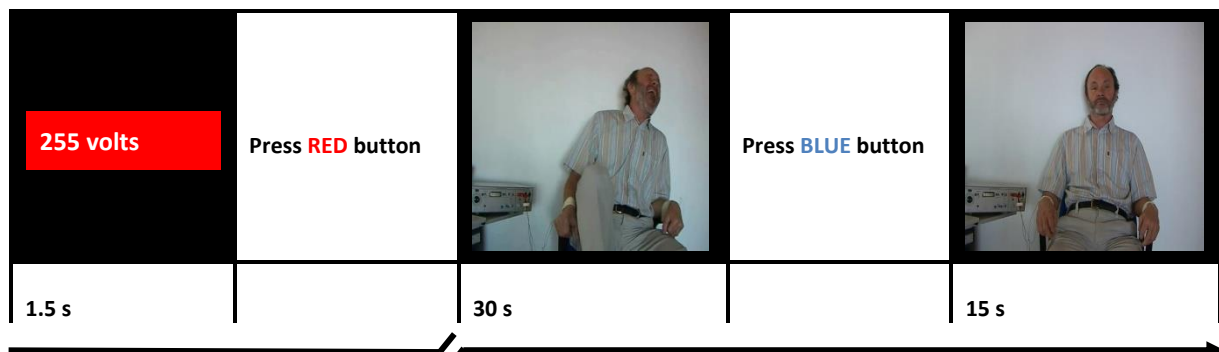
Then, participants were given a booklet about the purpose of the experiment (cover-story) and the experimental procedure. First, participants were informed that they were involved in an experiment on the effects of defibrillation on human memory; it was claimed that very little is known about the short-term effects of defibrillation on memory and how much these effects could bias medical diagnosis (i.e., misattribution of memory impairment). Participants were informed that we were bringing together a number of adults; we were asking some of them to pass a memory test and to receive electric shocks during this test (the “learners”), and some of them to administrate the shocks (the participants). The “learners” and the participants were supposed to be installed into two adjacent rooms. A video-conferencing system was supposed to ensure the communication. Then, participants were informed about the experimental procedure. The memory test was a paired-associate learning task. Before the experiments, the “learners” were supposed to memorize a series of word pairs; during the experiment, they were to indicate for each pair, which of four words had originally been paired with the first word. It was emphasized that to ensure the safety of the procedure (i) the memory test was administrated by a doctor, (ii) the shocks were delivered on the arms (and not on the chest), (iii) “learners” were strapped on a chair to avoid any dangerous movements. The notice provided very simple instructions to the participants: they were asked to follow the instructions appearing on the screen during the experiment (“press red button” to deliver the shock, “press blue button” to go to the next question). The last part provided Figure 1(a) with

the following explanations: (i) the experimental procedure was divided into six blocks, (ii) participants were asked to administer a shock before each question into four blocks (see Figure 1). It took approximately five minutes to the participants to read and memorize the booklet content. After having read the booklet, participants were invited to ask any questions about the procedure. Then, they gave their written consent for the participation.

The IVMOE task was adapted from Dambrun & Vatiné, (2010). The procedure started with the seating of the student on a chair by the virtual doctor (35 years-old male in a white technician's coat – 1.25 min): the learner was strapped into the chair and an electrode was attached on each arm. The next phase was the experimental procedure. In our immersive obedience paradigm, each trial begins by indicating the shock intensity (e.g., 120V) to be administered to the learner (see Figure 1(b)). At the level of 120V, the learner exhibited some cues of pain-related behavior (moan, facial expression of pain and corresponding body movement). At the level 180V, the learner's pain became more intense. At the level of 255V, the learner reacted strongly and asked to stop the experiment because the shocks were too painful. At the level of 300V, the learner's groaning was loud and asked to stop the experiment.



(a)



(b)

Figure 1. Experimental procedure. (a) Overview of the experimental procedure. The experimental procedure was structured into six blocks, with five trials (i.e. five questions to the learner) in the first five blocks, and three trials in the last block. The total experimental time was approximately 40 min. (b) Overview of the timeline of a pain trial. Each trial begins by indicating the shock intensity (e.g., 120V) to be administered to the learner. In the pain trials, participants had to press the red button and the subsequent suffering of the learner was observed. Each shock is accompanied by an “electric” buzz sound. Then participants had to press the blue button. The words were enunciated by the virtual doctor and the learner gave his answer. In the non-pain trials, participants had to press the blue button and the learner was simply observed giving his answer (30 s). All parameters were computer-programmed in order to give the impression of real time and a dynamic environment.

When participants expressed their desire to stop the experiment, the experimenter intervened with four consecutive injunctions: (i) “Please continue”, (ii) “It is important for the experiment that you continue”, (iii) “The procedure requires that you continue”, (iv) “You have no other choice, you must go on”. If participants insisted on stopping the experiment after the fourth injunction, the experiment was terminated. Otherwise, this four-part sequence was restarted each time the participants asked whether they could stop the experiment.

Obedience was measured as follow: (i) absolute obedience was indicated by whether participants delivered all the shocks (0 = *not fully obedient*, 1 = *fully obedient*), (ii) the relative obedience refers to the maximum number of shocks administrated (from 0 to 18). Participants were fully debriefed after the experiment about their experience during the obedience procedure with particular reference to the impact of observing the learner in pain.

Data analysis

Prior to analysis, signal was visually inspected and ectopic (abnormal) heart beats were removed using the LabChart Beat Classifier view. The HRV 2.0 Add-On for LabChart was used to calculate RMSSD (root mean square of successive R-R differences). As mentioned previously, past research demonstrates that increased RMSSD is a biological marker of mental effort. Moreover, this measure of HRV has been shown to strongly relate to high frequency power bands in spectral analyses (e.g., Friedman, Allen, Christie, & Santucci, 2002), providing evidence for its use as an appropriate index of vagal tone.

Effects of resting HRV (baseline 1) and RWA were tested using multiple regression analysis. Behavioral consequences of experimental manipulations were tested using multiple regression analysis, with dummy codes for experimental condition. Continuous predictors were centered before analysis. Subjective consequences of experimental manipulations were tested using analysis of variance. The physiological consequences of experimental manipulations were tested using repeated measures analysis of variance. Within-subjects analyses employed the Greenhouse-Geisser correction to degrees of freedom for departures from sphericity as indicated by epsilon values. Because destructive obedience is best represented as a binary response variable (i.e. as evidenced by a bimodal distribution of scores) rather than a continuous variable, we used the binary score of obedience for our

analysis. However, analyses with relative obedience as a dependent variable are given in Appendix A and Appendix B.

Results

Rates of obedience

The distribution of break-off points is presented in table 1. On the whole, 51.5% of participants fully obeyed (42% in the “high SRE” condition and 61% in the “low SRE” condition). The mean shock level was 12.10 (10.66 in the “high SRE” condition and 13.57 in the “low SRE” condition).

Table 1

Distribution of break-off points

Shock level	Verbal designation and voltage level	Number of participants for whom this was maximum shock		
		Overall (<i>n</i> = 99)	Low SRE (<i>n</i> = 49)	High SRE (<i>n</i> = 50)
	Refuse	9	3	6
	No shock			
	0			
	0			
	0			
	0			
	0	8	1	7
	Moderate shock			
1	120	3	2	1
2	120			1
3	120			
4	120	2	2	
5	120	3	1	2
	Painful shock			
6	180	3	3	
7	180			
8	180			
9	180			
10	180	3	1	2
	No shock			
	0			
	0			
	0			
	0			
	0			
	Intense shock			
11	255	4		4
12	255	8	4	4
13	255			
14	255	1		1
15	255			
	Extreme intensity shock			
16	300	1		1
17	300	2	2	
18	300	51	30	21
	Percentage of obedience	51.50%	61%	42%
	Mean shock level (SD)	12.10 (7.29)	13.57 (6.64)	10.66 (7.67)

Note. SRE = self-regulatory effort. SD = standard deviation.

Data description

Descriptive statistics are displayed for all conditions (figure 2), low SRE condition (figure 3) and high SRE condition (figure 4).

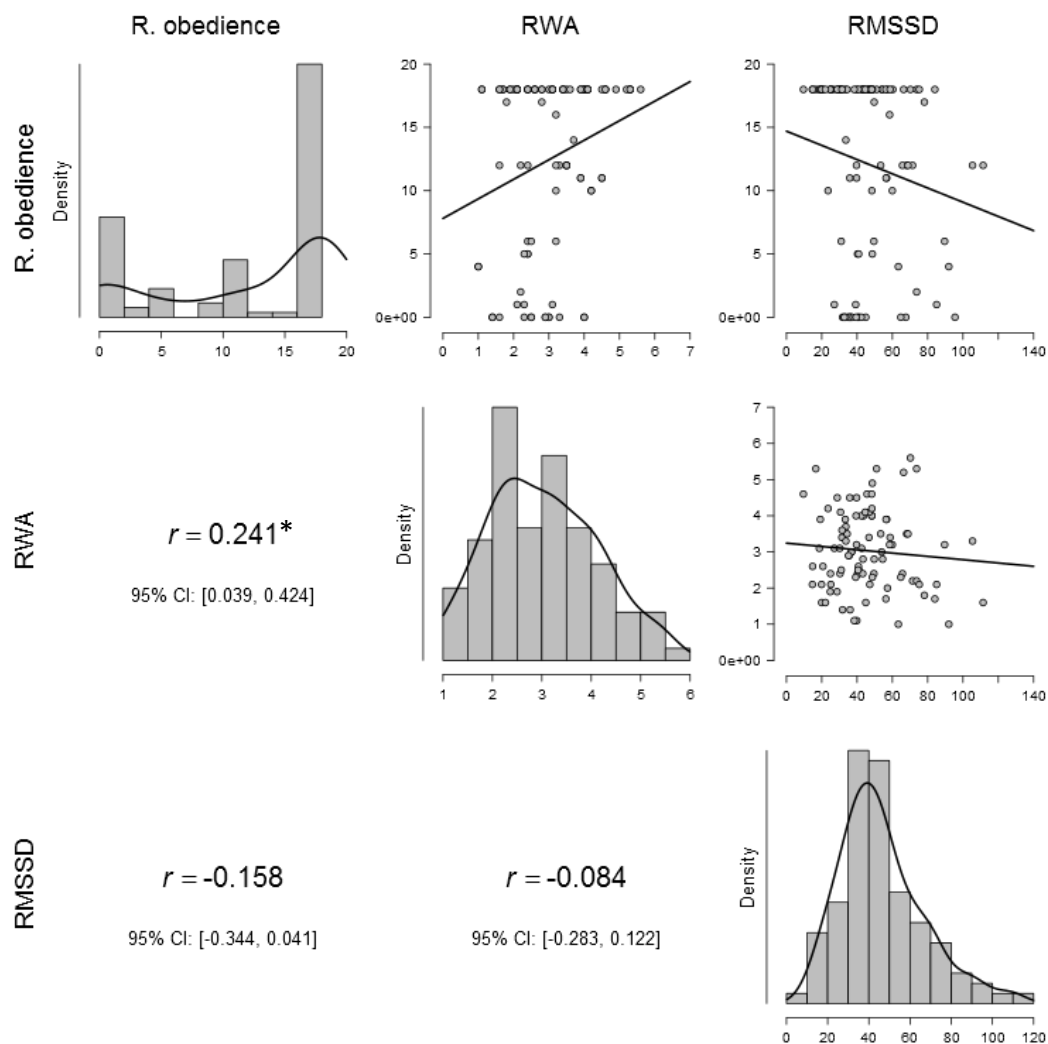


Figure 2. Scatter plots, densities, and Pearson correlation coefficients between recorded variables (**overall**). A. obedience = absolute obedience; R. obedience = relative obedience; RWA = right-wing authoritarianism; RMSSD = root mean squared successive differences in the interbeat interval (baseline 1). * $p < .05$

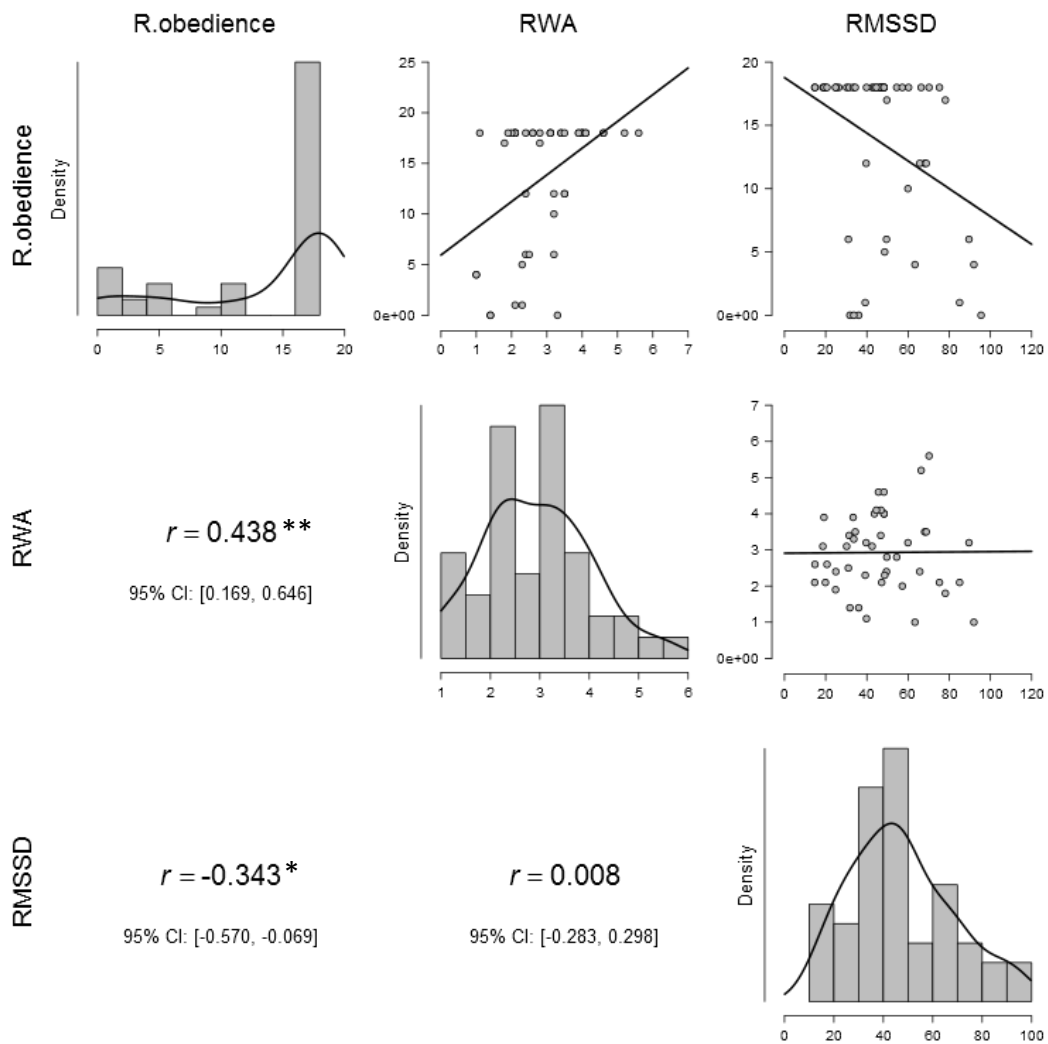


Figure 3. Scatter plots, densities, and Pearson correlation coefficients between recorded variables (**low SRE condition**). A. obedience = absolute obedience; R. obedience = relative obedience; RWA = right-wing authoritarianism; RMSSD = root mean squared successive differences in the interbeat interval (baseline 1). * $p < .05$, ** $p < .01$

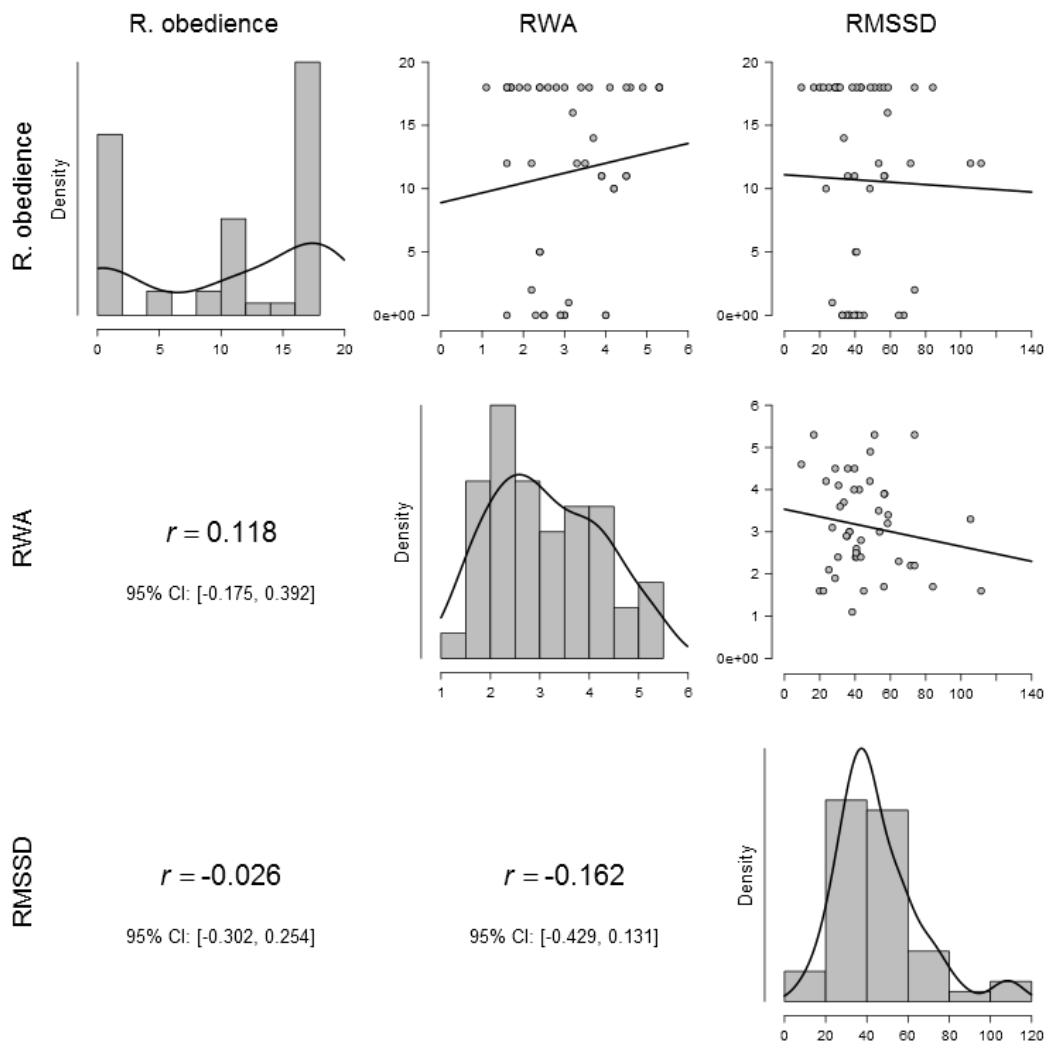


Figure 4. Scatter plots, densities, and Pearson correlation coefficients between recorded variables (**high SRE condition**). A. obedience = absolute obedience; R. obedience = relative obedience; RWA = right-wing authoritarianism; RMSSD = root mean squared successive differences in the interbeat interval (baseline 1).

Predicting Obedience to Authority

As illustrated in Figure 5, higher HRV (baseline 1) predicted more disobedience, $OR = 1.64$, $SE = .27$, $W = 6.33$, $p = .01$, $95\% CI [.29, .86]$. Replicating previous findings, we also found that higher RWA scores predicted more obedience, $OR = 5.31$, $SE = .25$, $W = 4.15$, $p = .04$, $95\% CI [1.02, 2.75]$.

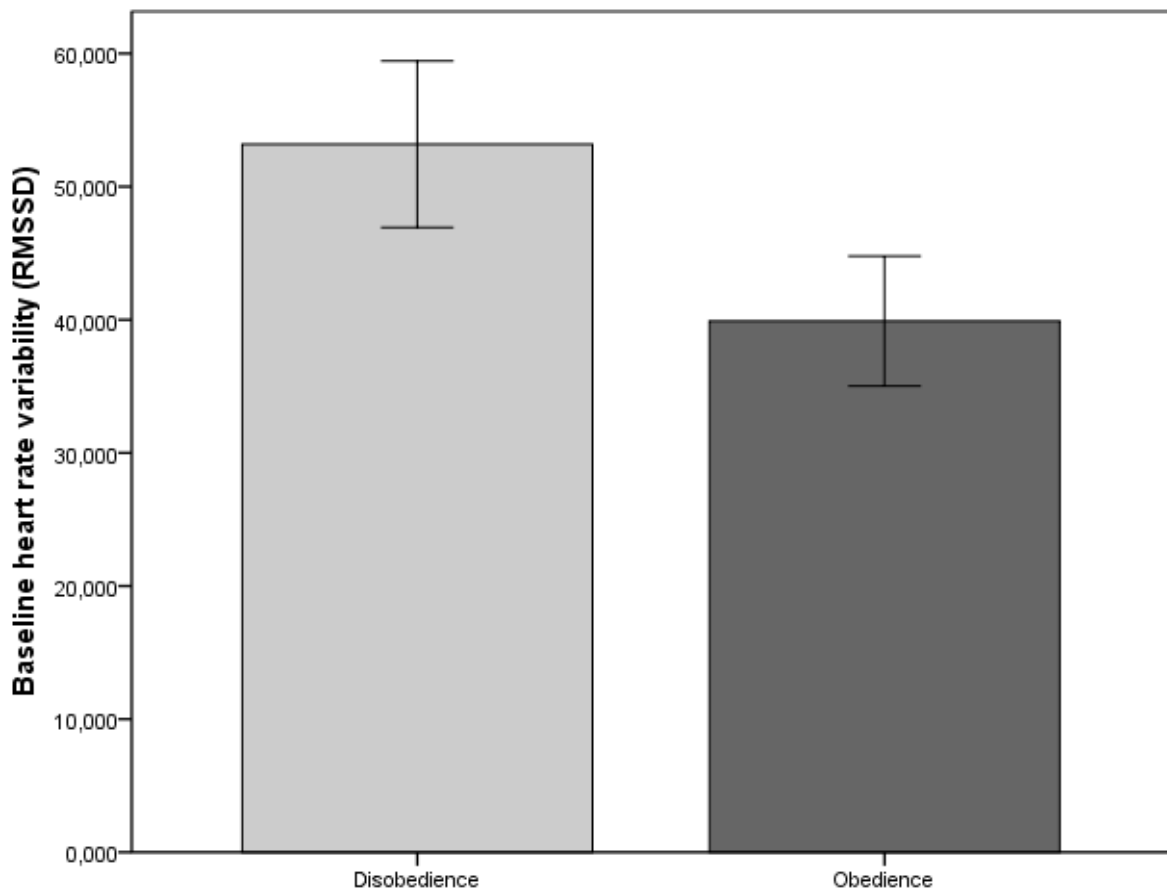


Figure 5. Heart rate variability (baseline 1) as a function of absolute obedience. Mean changes are shown with 95% CI. RMSSD = root mean squared successive differences in the interbeat interval.

Subjective and Physiological Consequences of Self-Regulatory Effort

The fatiguing task was perceived as being significantly more difficult than the control task ($M = 5.50$, $SD = 1.54$, vs. $M = 3.81$, $SD = 1.74$, on a scale from 1 to 8), $F(1, 77) = 20.76$, $p = .000$, $\eta_p^2 = .21$, 95% CI [-2.42, -.95]. HRV was significantly increased after high SRE relative to low effort, as reflected in a significant interaction of time (baseline 1 vs. baseline 2) and self-regulation condition (high vs. low effort), $F(1, 95) = 11.72$, $p = .001$, $\eta_p^2 = .11$.

Figure 6 shows HRV changes from baseline.

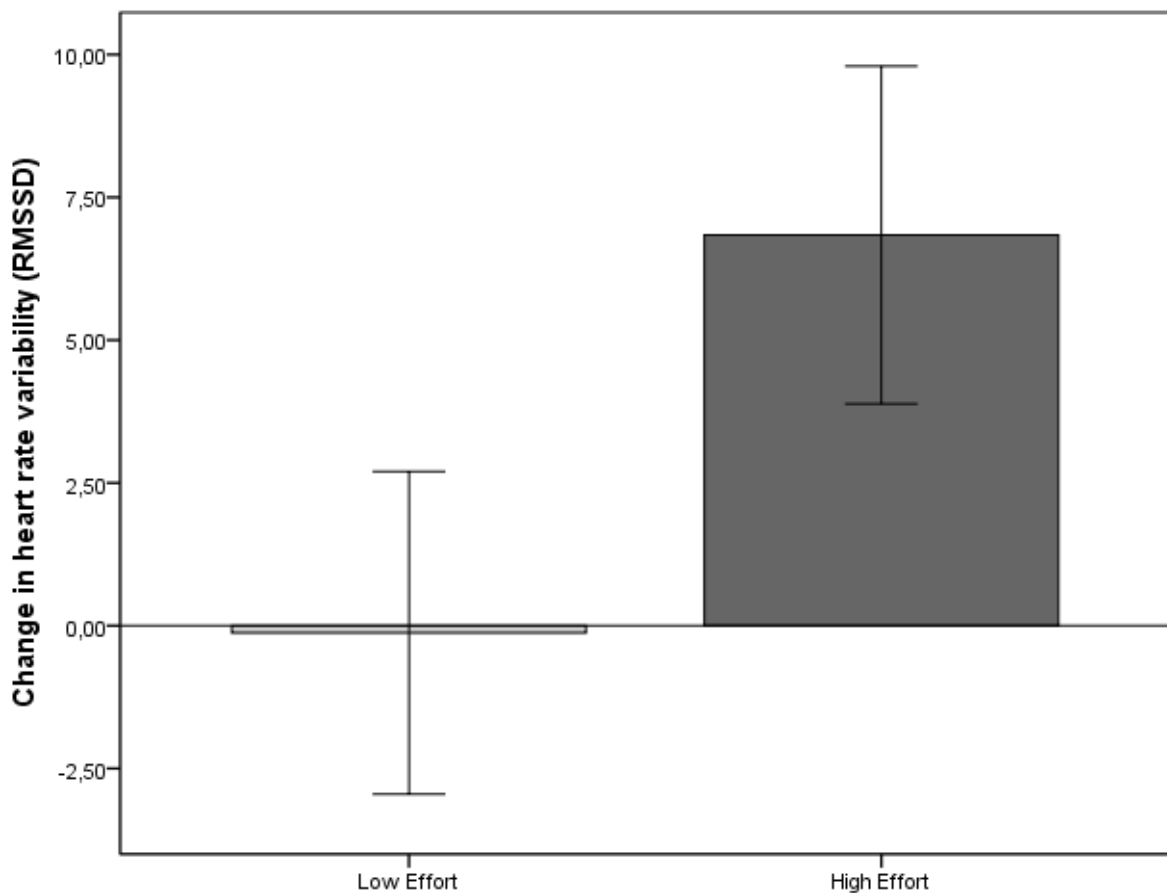


Figure 6. Change in heart rate variability as a function of self-regulation condition (baseline 2 – baseline 1). Mean changes are shown with 95% CI. RMSSD = root mean squared successive differences in the interbeat interval.

Behavioral Consequences of Self-Regulatory Effort

SRE affected subsequent obedience behavior, as reflected in a significant interaction between RWA and self-regulation condition, $OR = 1.82$, $SE = .24$, $W = 4.05$, $p = .04$, 95% CI [.37, .98]. It seems that the effect of RWA was absent among participants who had previously exercised high SRE (see figure 7 and table 2). The main effect of RWA was significant, $OR = 5.15$, $SE = .24$, $W = 4.10$, $p = .04$, 95% CI [1.01, 2.67]. We also found a marginally significant main effect of self-regulation condition, $OR = 1.91$, $SE = .22$, $W = 3.49$, $p = .06$, 95% CI [.41, 1.02].

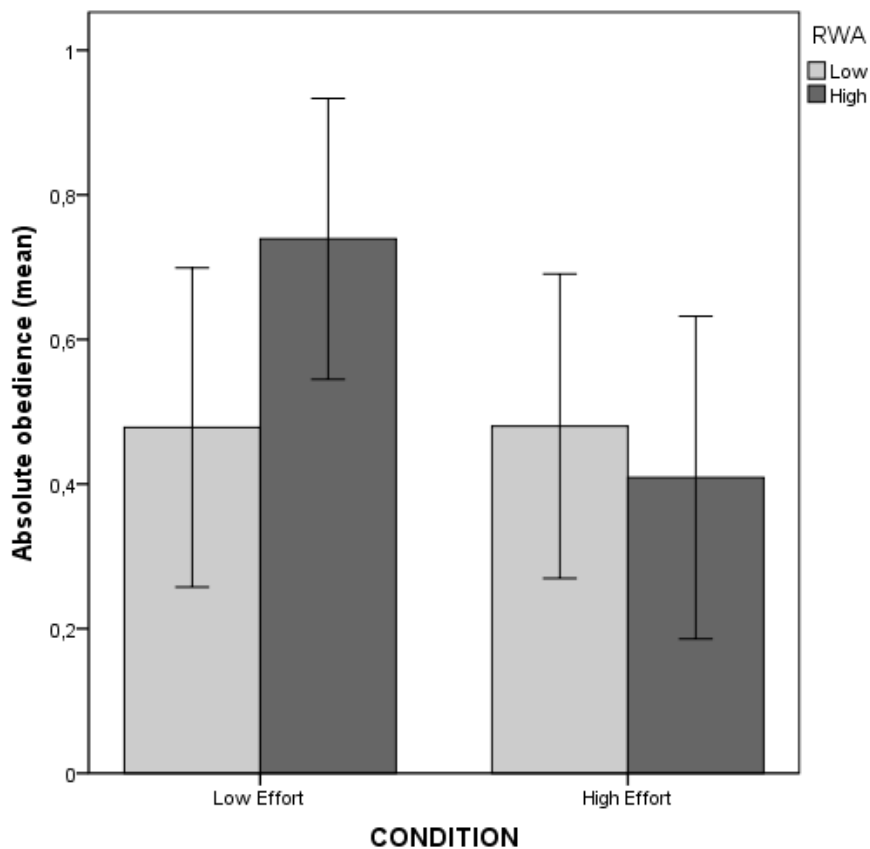


Figure 7. Absolute obedience as a function of self-regulation condition and RWA scores (median split). Mean changes are shown with 95% CI. RWA = right-wing authoritarianism.

Table 2

Relative obedience as a function of self-regulation condition and RWA scores (median split)

	High RWA	Low RWA	<i>t</i> test for difference		
			<i>t</i>	<i>p</i>	95% CI
High SRE	12.77 (5.98)	10.08 (8.39)	-1.25	.21	[-7.03, 1.64]
Low SRE	15.57 (4.78)	11.78 (7.45)	-2.04	.04	[-7.50, -.05]

Note. Standard deviations are in parentheses. RWA = right-wing authoritarianism; SRE = self-regulatory effort.

Discussion

The present study aimed to advance our understanding of the role of self-regulation processes in the Milgram paradigm. In order to do so, we examined the influence of resting HRV and SRF on destructive obedience. First, we found that low-RWA individuals and high-HRV individuals were less obedient. Second, our results suggest that low self-regulatory inhibition (induced by fatigue) reduced destructive obedience. Finally, we found that high-RWA individuals were more obedient only when *not* experiencing SRF.

Heart rate variability

The influence of HRV on destructive obedience is consistent with specific predictions of the polyvagal theory (Porges, 2007). This phylogenetic theory of adaptive reactions to challenge highlights that physiological states support different classes of behavior: (i) a physiological state characterized by a vagal withdrawal would support the mobilization of fight and flight behaviors, while (ii) a physiological state characterized by increased vagal influence on the heart would support spontaneous social engagement behaviors. Following

this perspective, prosocial behaviors are underpinned by vagal tone and individuals with greater cardiac vagal tone should exhibit a greater range of prosocial behaviors. However, very few studies used behavioral measures to establish this relationship between vagal tone and prosociality (Beffara, Bret, Vermeulen, & Mermillod, 2016). Our results support that vagal functioning may contribute to prosocial disobedience. In our view, such a physiological signature constitutes an important construct regardless of whether its causes are genetic, environmental, or a combination of both. And as emphasized by Blascovich (2014), the advantages of psychophysiological method “accrue from its stipulation of relatively unbiased, real-time indexes of psychological processes related to motivational, affective, and cognitive constructs – ones that are often difficult to quantify without bias or artifact in the absence of such objective assessments and impossible to quantify, especially temporally, otherwise in vivo.”

Self-regulatory fatigue

Prior research demonstrates that exerting high SRE at Time 1 impairs self-control at Time 2. It seems to be a widely-held assumption that SRF decreases the likelihood of prosocial behavior by enhancing the influence of automatic emotional processes (e.g., Mead, Baumeister, Gino, Schweitzer, & Ariely, 2009). But growing evidence supports the notion that human emphatic and sympathetic inclinations are based on pre-programmed, neuro-endocrine and emotional feedbacks operating as “gut-feelings” (e.g., Høgh-Olesen, 2010). It has been proposed that impulsive aggression arises as a consequence of faulty regulation of negation emotion (e.g., Davidson, Putnam, & Larson, 2000), but harmful behavior may also result from *impaired aversive responses* to the others’ distress (e.g., Blair, 2007). According to the “shifting priorities model”, (i) self-control failure due to initial SRE relies on a motivation-shift mechanism, (ii) automatic impulse is not inherently “good” or “bad” (e.g.,

Inzlicht et al., 2014). In our study, one could hypothesize that shifts in motivation drove the participants' attention on pain cues from the victim and favored a more accurate inference of the internal experience of the victim, thereby increasing the value of giving in to the automatic impulse. A recent meta-analysis showed that disobedience in the Milgram paradigm is most likely to the point at which the victim first request to be released (Packer, 2008). Interestingly, the distribution of break-off points by condition in our study revealed a similar critical decision point in the high SRE condition (i.e., increased disobedience in block 5) but not in the low SRE condition. This result suggests that SRF increased disobedience during the periods in which the victim's protests became more vehement and emotional.

Self-regulatory fatigue by right-wing authoritarianism interaction

Across human cultures, RWA appears to be a very stable political attitude over the lifecourse (e.g. Hatemi & McDermott, 2012). A great deal of research has been devoted to explaining the origins and manifestations of ideology. Studies found that making threats salient makes people more inclined towards right-wing beliefs (e.g., Duckitt, 2013). According to Duckitt and colleagues, RWA expresses the motivational goal of establishing or maintaining collective security (e.g., Duckitt & Sibley, 2010). Studies found that individuals for whom stereotypes are automatically activated are more likely to display prejudice and engage in discriminatory behavior after SRE (e.g., Govorun & Payne, 2006). Therefore, one could expect that SRF increases the influence of RWA on destructive obedience. Linking research on attitudes to contemporary models of self-regulation, studies investigated the behavioral impact of mental fatigue and automatic attitudes versus personal standards. Automatic attitudes refer to spontaneous evaluations that have their roots in associative processes of spreading activation (Strack & Deutsch, 2004). Personal standards have their origin in reflective processes of higher-order propositional reasoning (e.g., Hofmann, Rauch,

& Gawronski, 2007). As emphasized by Hofmann et al. (2007), “impulsive action tendencies resulting from automatic evaluations are often in conflict with deliberate action tendencies resulting from personal goals or standards, implying a tug-of-war similar to the one proposed by contemporary models of self-regulation”. Because reflective processes are deliberate and effortful, the authors hypothesized and found evidence that SRF decreases the behavioral impact of personal standards.

All in all, our findings suggest (i) that stress vulnerability predicted destructive obedience, (ii) that self-regulatory impairment was responsible for decreased destructive obedience, (iii) that RWA predicted destructive obedience when *not* combined with SRF.

Conclusion

Cheetham et al. (2009) found that the observation of the victim’s pain during the obedience procedure induced brain activations related to an aversive state of personal distress, and our findings provide evidence that inter-individual differences in stress vulnerability (as indexed by resting HRV) underpinned inter-individual differences in destructive obedience. There is a growing appreciation that distressed individuals tend to inhibit pain-related stress responses (e.g., Goubert et al., 2009; Petrovic et al., 2004) and that antisocial behavior may result from low automatic aversive responses to the others’ pain (e.g., Blair, 2007). We found evidence that low self-regulatory control favored disobedience toward destructive orders and suppressed the well-established influence of RWA. Our results support the notion that inhibitory control is needed to comply with destructive orders, and that prosocial disobedience could be more easily grasped relying on automatic processes.

In accordance with a “Western dualistic tradition that pits body against mind and culture against instinct” (de Waal, 2003), current theories and models of destructive

obedience emphasis on cognition/rationality and downplay emotions as guides of human behavior (e.g., Bandura, 1999; Reicher & Haslam, 2011; Milgram, 1974; Kelman & Hamilton, 1989). The present study throws some light on the physiological and emotional underpinnings of destructive obedience.

References

- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Aldao, A., Dixon-Gordon, K. L., & De Los Reyes, A. (2015). Individual differences in physiological flexibility predict spontaneous avoidance. *Cognition and Emotion*, *30*(5), 985-998. doi:10.1080/02699931.2015.1042837
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Perseus Books.
- Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Canada: University of Manitoba Press.
- Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.
- Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, *10*(3), 229-240. doi:10.1037/1089-2680.10.3.229
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193-209. doi:10.1207/s15327957pspr0303_3
- Baumeister, R. F., Sparks, E. A., Stillman, T. F., & Vohs, K. D. (2008). Free will in consumer behavior: Self-control, ego depletion, and choice. *Journal of Consumer Psychology*, *18*(1), 4-13. doi:10.1016/j.jcps.2007.10.002
- Beauvois, J. L., Courbet, D., & Oberlé, D. (2012). The prescriptive power of the television host: A transposition of Milgram's obedience paradigm to the context of TV game show. *European Review of Applied Psychology*, *62*(4), 111-119. doi:10.1016/j.erap.2012.02.001

- Beffara, B., Bret, A.G., Vermeulen, N., & Mermillod, M. (2016). Resting High Frequency Heart Rate Variability Selectively Predicts Cooperative Behavior. *Physiology & Behavior*, *164*, 417-428. doi:10.1016/j.physbeh.2016.06.011
- Bègue, L., Beauvois, J. L., Courbet, D., Oberlé, D., Lepage, J., & Duke, A. (2015). Personality Predicts Obedience in a Milgram Paradigm. *Journal of Personality*, *83*(3), 299-306. doi:10.1111/jopy.12104
- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*(9), 387-392. doi:10.1016/j.tics.2007.07.003
- Blascovich, J. (2014). Using Physiological Indexes in Social Psychological Research. In H. T., Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd Ed., pp. 101-122). New York: Cambridge University Press.
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychology Inquiry*, *13*(2), 103-124. doi:10.1207/S15327965PLI1302_01
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, *60*(3), 398-413. Retrieved from <http://www.stanleymilgram.com/pdf/understanding20%behaviour.pdf>
- Brosschot, J. F., Verkuil, B., & Thayer, J. F. (2016). The default response to uncertainty and the importance of perceived safety in anxiety and stress: an evolution-theoretical perspective. *Journal of Anxiety Disorders*, *41*, 22-34. doi:10.1016/j.janxdis.2016.04.012.

- Cheetham, M., Pedroni, A. F., Antley, A., Slater, M., & Jäncke, L. (2009). Virtual milgram: empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Frontiers in Human Neuroscience*, 3(29). doi:10.3389/neuro.09.029.2009
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, 107(40), 17433-17438. doi:10.1073/pnas.1009396107
- Dang, J. (2017). An updated meta-analysis of the ego-depletion effect. *Psychological Research*, 1-7. doi:10.1007/s00426-017-0862-x
- Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, 41(2), 81-94. doi:0166-4328(90)90144-4
- Dambrun, M., & Vatiné, E. (2010). Reopening the study of extreme social behaviors: Obedience to authority within an immersive video environment. *European Journal of Social Psychology*, 40(5), 760-773. doi:10.1002/ejsp.646
- Davidson, R. J., Putnam, K. M., & Larson, C. L. (2000). Dysfunction in the neural circuitry of emotion regulation – a possible prelude to violence. *Science*, 289(5479), 591-594. doi:10.1126/science.289.5479.591
- Decety, J., & Lamm, C. (2009). Empathy versus Personal Distress: Recent Evidence from Social Neuroscience. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 199-213). Cambridge, MA: MIT Press.
- Decety, J., & Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience*, 2(1), 1-24. doi:10.1016/j.dcn.2011.05.003

- de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Science*, 10(10), 435-441. doi:10.1016/j.tics.2006.08.008
- de Waal, F. B. M. (2003). Morality and the Social Instincts: Continuity with the Other Primates. In G. B. Peterson (Ed.), *The Tanner Lectures on Human Values* (pp. 1-40). Salt Lake City, UT: University of Utah Press. Retrieved from http://tannerlectures.utah.edu/_documents/a-to-z/d/deWaal_2005.pdf
- Duckitt, J. (2013). Introduction to the special section on authoritarianism in societal context: The role of threat. *International Journal of Psychology*, 48(1), 1-5. doi:10.1080/00207594.2012.738298
- Duckitt, J., & Sibley, C. G. (2010). Personality, ideology, prejudice, and politics: A dual process motivational model. *Journal of Personality*, 78(6), 1861-1894. doi:10.1111/j.1467-6494.2010.00672.x
- Elms, A. C. (1995). Obedience in Retrospect. *Journal of Social Issues*, 51(3), 21-31. doi:10.1111/j.1540-4560.1995.tb01332.x
- Elms, A. C., & Milgram, S. (1966). Personality characteristics associated with obedience and defiance toward authoritative command. *Journal of Experimental Research in Personality*, 1(4), 282-289.
- Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71-83). Cambridge, MA: MIT Press.
- Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P., et al. (1994). The relations of emotionality and regulation to dispositional and situational empathy-related responding. *Journal of Personality and Social Psychology*, 66(4), 776-797.

- Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg & W. Damon (Eds.), *Handbook of child psychology. Social, emotional, and personality development* (Vol. 3, 6th ed., pp. 646-718.). New York: Wiley.
- Fabes, R. A., Eisenberg, N., & Eisenbud, L. (1993). Behavioral and physiological correlates of children's reactions to others' distress. *Developmental Psychology*, 29(4), 655-663. doi:10.1037/0012-1649.29.4.655
- Feldman, S., Lavine, H., Lodge, M., & Verhulst, B. (2010). *Seeing Negative: Authoritarianism and Automatic Vigilance for Threatening Stimuli*. Paper presented at the Annual Meeting of the International Society of Political Psychology, San Francisco, CA.
- Friedman, B. H., Allen, M. T., Christie, I. C., & Santucci, A. K. (2002). Validity concerns of common heart rate variability indices. Addressing quantification issues in time- and frequency-domain measures of HRV. *IEEE Engineering in Medicine and Biology Magazine*, 21, 35-40. doi:10.1109/MEMB.2002.1032637
- Goubert, L., Craig, K. D., & Buysse, A. (2009). Perceiving Others in Pain: Experimental and Clinical Evidence on the Role of Empathy. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 153-166). Cambridge, MA: MIT Press.
- Govorun, O., & Payne, K. B. (2006). Ego-Depletion and Prejudice: Separating Automatic and Controlled Components. *Social Cognition*, 24(2), 111-136. doi:10.1521/soco.2006.24.2.111
- Gross, J. J. (2002). Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology*, 39(3), 281-91. doi:10.1017.S0048577201393198
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implication for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348-362. doi:10.1037/0022-3514.85.2.348

- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*(1), 98-116. doi:0.1007/s11211-007-0034-z
- Hatemi, P. K., & McDermott, R. (2012). The genetics of politics: discovery, challenges, and progress. *Trends in Genetics, 28*(10), 525-533. doi:10.1016/j.tig.2012.07.004
- Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology, 43*(3), 497-504. doi:10.1016/j.jesp.2006.05.004
- Høgh-Olesen, H. (2010). Homo sapiens - Homo socius: A comparative analysis of human mind and kind. In H. Høgh-Olesen (Ed.), *Human morality and sociality: Evolutionary and comparative perspectives* (pp. 235-271). Basingstoke: Palgrave Macmillan.
- Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why Self-Control Seems (But May Not Be) Limited. *Trends in Cognitive Sciences, 18*(3), 127-133. doi:10.1016/j.tics.2013.12.009.
- Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of Obedience: Toward a Social Psychology of Authority and Responsibility*. New Haven: Yale University Press.
- Kool, W., & Botvinick, M. (2014). A labor-leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General, 143*(1), 131-141. doi:10.1037/a0031048.
- Lankford, A. (2009). Promoting aggression and violence at Abu Ghraib: The U.S. military's transformation of ordinary people into torturers. *Aggression and Violent Behavior, 14*(5), 388-395. doi:10.1016/j.avb.2009.06.007
- Mead, N., Baumeister, R. F., Gino, F., Schweitzer, M., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology, 45*(3), 594-597. doi:10.1016/j.jesp.2009.02.004

- Mermillod, M., Marchand, V., Lepage, J., Bègue, L., & Dambrun, M. (2015). Destructive Obedience Without Pressure: Beyond the Limits of the Agentic State. *Social Psychology, 46*(6), 345-351. doi:10.1027/1864-9335/a000251
- Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371-378. doi:10.1037/h0040525
- Milgram, S. (1965). Some Conditions of Obedience and Disobedience to Authority. *Human Relations, 18*(1), 57-76. doi:10.1177/001872676501800105
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. London: Tavistock Publications.
- Milyavskaya, M., & Inzlicht, M. (2016). Attentional and motivational mechanisms of self-control. In D. de Ridder, M. Adriaanse, & K. Fujita (Eds), *Handbook of Self-Control in Health and Well-Being*. New York: Routledge. Retrieved from <https://static1.squarespace.com/static/550b09eae4b0147d03eda40d/t/56f9cb3907eaa0601bad7717/1459211067303/attentional-and-motivational-mechanisms-of-self-control.pdf>
- Movius, H. L., & Allen, J. J. B. (2005). Cardiac vagal tone, defensiveness, and motivational style. *Biological Psychology, 68*(2), 147-162. doi:10.1016/j.biopsycho.2004.03.019
- Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247-259. doi:10.1037/0033-2909.126.2.247
- Packer, D. J. (2008). Identifying systematic disobedience in Milgram's obedience experiments: A meta-analytic review. *Perspectives on Psychological Science, 3*(4), 301-304. doi:10.1111/j.1745-6924.2008.00080.x

- Petrovic, P., Carlsson, K., Petersson, K. M., Hansson, P., & Ingvar, M. (2004). Context-dependent deactivation of the amygdala during pain. *Journal of Cognitive Neuroscience*, *16*(7), 1289-1301. doi:10.1162/0898929041920469
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, *74*(2), 116-143. doi:10.1016/j.biopsycho.2006.06.009
- Quintana, D. S., & Heathers, J. A. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology*, *5*, 805. doi:10.3389/fpsyg.2014.00805
- Reicher, S. D., & Haslam, S. A. (2011). After shock? Towards a social identity explanation of the Milgram 'obedience' studies. *The British Journal of social psychology*, *50*(Pt 1), 163-169. doi:10.1111/j.2044-8309.2010.02015.x.
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609-612. doi:10.1016/j.jrp.2013.05.009
- Segerstrom, S. C., & Ness, S. L. (2007). Heart rate variability indexes self-regulatory strength, effort, and fatigue. *Psychological Science*, *18*(3), 275-281. doi:10.1111/j.1467-9280.2007.01888.x
- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *Plos One*, *1*(1), e39. doi:10.1371/journal.pone.0000039
- Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, *8*(3), 220-247. doi:10.1207/s15327957pspr0803_1
- Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability. Standards of measurement, physiological interpretation and clinical use. *Circulation*, *93*(5), 1043-1065. Retrieved

from http://www.escardio.org/static_file/Escardio/Guidelines/Scientific-Statements/guidelines-Heart-Rate-Variability-FT-1996.pdf

Footnote 1

We used the overall effect size found for mental fatigue in the Hagger et al. (2010) meta-analysis ($d = .62$). The specific effect size obtained for the crossing-out-letters task was higher ($d = .77$), but we preferred to use a more conservative estimator to maximize our chances of detecting an effect (i.e. by means of an appropriate number of participants). This higher number of participants also ensure more reliability and stability of the p-values reported in this experiment (Schönbrodt & Perugini, 2013).

Appendix A

Multiple regression examining the associations between relative obedience and the expected predicting variables (resting HRV and RWA)

Criterion variables	Predictor variables	β	t	p	95% CI
Relative obedience	HRV	-.03	-.33	.73	[-1.75, 1.24]
	RWA	.26	2.66	.03	[.16, 3.28]

Note. HRV = heart rate variability; RWA = right-wing authoritarianism.

Appendix B

Multiple regression examining the behavioral consequences of SRE

Criterion variables	Predictor variables	β	t	p	95% CI
Relative obedience	RWA	.26	2.66	.009	[.47, 3.26]
	SRE	-.19	-1.92	.05	[-2.73, .04]
	SRE*RWA	-.14	-1.44	.15	[-2.40, .37]

Note. RWA = right-wing authoritarianism; SRE = self-regulatory effort.

Influence de l'autoritarisme, du tonus vagal et de la fatigue mentale sur la soumission à l'autorité

Comme pour l'étude 3, nous trouvons dans l'étude 4 une influence du tonus vagal sur l'OD. Les résultats de cette étude montrent également que l'induction d'une fatigue mentale tend à diminuer l'OD et supprime l'influence du RWA sur l'OD. Une certaine tradition de recherche en psychologie propose que la fatigue mentale diminue la probabilité d'adopter un comportement prosocial en augmentant l'influence des processus émotionnels automatiques (e.g., Mead, Baumeister, Gino, Schweitzer, & Ariely, 2009). Mais un nombre important de travaux montrent que l'empathie et la sympathie reposent sur des processus émotionnels automatiques sous-tendus par des structures cérébrales anciennes du point de vue évolutif (e.g., Decety & Svetlova, 2012). La diminution observée de l'influence du RWA sur l'OD après induction d'une fatigue mentale est cohérente avec les travaux explorant les effets comportementaux de la fatigue mentale et des attitudes automatiques (*i.e.*, évaluations spontanées reposant sur des mécanismes associatifs ; Strack & Deutsch, 2004) *vs* standards personnels (reposant sur des processus réflexifs ; Hofmann, Rauch, & Gawronski, 2007). Des études ont en effet montré que l'induction d'une fatigue mentale diminue l'impact comportemental des standards personnels (e.g., Hofmann et al., 2007).

L'effet d'interaction entre RWA et fatigue mentale suggère que les personnes présentant de forts scores de RWA sont plus obéissantes à condition de pouvoir exercer un effort cognitif. Dans leur ensemble, les données de l'étude 4 supportent l'hypothèse voulant que l'exercice d'un contrôle inhibiteur sur la résonance empathique augmente l'OD, alors que les réponses émotionnelles automatiques favorisent la désobéissance à des ordres destructeurs.

PARTIE 5. SUBSTRATS NEURONAUX DE L'OBEISSANCE DESTRUCTRICE

5.1. Étude EEG : Obéissance destructrice et contrôle inhibiteur

Objectifs et présentation de l'étude

L'objectif de cette étude était de mettre en relation l'OD avec l'activité électroencéphalographique des participants. Plus précisément, nous nous sommes intéressés aux oscillations neurales dans la gamme de fréquence thêta du fait de leur implication dans le contrôle inhibiteur des réponses émotionnelles aversives (e.g., Ertl et al., 2013). Nous nous attendions à ce qu'une augmentation de la puissance thêta soit associée à l'OD. Nous nous sommes également intéressé aux oscillations neurales dans la gamme de fréquence alpha (8–12Hz) du fait de leur implication dans l'attention portée aux stimuli externes (e.g., Pfurtscheller & Lopes da Silva, 1999). Dans la partie introductive de ce travail doctoral, nous avons mentionné le fait que l'IVMOE est un outil adapté à l'étude expérimentale de l'OD. Cependant, d'aucun pourrait affirmer que des différences interindividuelles dans la « présence » pourraient sous-tendre les différences interindividuelles dans l'OD. Plus précisément, une moindre « présence » pourrait diminuer l'attention et l'implication des participants et ainsi augmenter leur propension à obéir. Des études EEG ont montré que l'amplitude des ondes alpha est inversement reliée aux processus attentionnels (e.g., Cook et al., 1999 ; Laufs et al., 2003). Ainsi, nous pourrions observer une augmentation de la puissance alpha chez les participants obéissants (vs désobéissants). Afin de tester nos hypothèses, nous avons enregistré l'activité électroencéphalographique des participants ($n = 50$) durant la procédure d'obéissance.

Manuscrit 5¹⁰

¹⁰ Manuscrit soumis pour publication auprès de la revue *Biological Psychology*.

Exploring Neural Signatures of Obedience to Authority

Johan Lepage¹, Laurent Bègue¹, Sylvain Harquel^{2,3} & Martial Mermillod^{2,4}

¹ Laboratoire InterUniversitaire de Psychologie, Université Grenoble Alpes, Grenoble, France

² Laboratoire de Psychologie et Neurocognition, CNRS UMR 5105, Université Grenoble Alpes, Grenoble, France

³ CNRS, INSERM, Université Grenoble Alpes, CHU Grenoble, IRMaGe, Grenoble, France

⁴ Institut Universitaire de France, Paris, France

Corresponding author

Johan Lepage, PhD Candidate: Laboratoire InterUniversitaire de Psychologie, UFR SHS, 1251 avenue Centrale, BP 47, 38040 Grenoble Cedex 9, France; E-mail: johan.lepage@gmail.com, Phone: (+33 4) 76 82 58 92, Fax: (+33 4) 76 82 56 65

Abstract

The mechanisms responsible for destructive obedience remain unknown. However, recent research suggests that the observation of the victim's pain within the Milgram paradigm induced neural activations related to a self-oriented aversive state of personal distress. Studies suggest that distressed individuals tend to inhibit their affective responses toward the others' pain in an effort to reduce their own discomfort. This strategy could have detrimental effects, such as decreased harm aversion. In the present study, we hypothesized that obedient participants would exert an inhibitory control over their emotional responses to the victim's pain. Recent findings suggest that theta oscillations are a key physiological mechanism of down-regulation of negative emotions. Studies also found that alpha oscillations are inversely related to attention and involvement. Whereas we didn't find any differences between obedient and disobedient participants in alpha power, on the overall experience, we predicted and found higher theta power in large areas among obedient individuals. These results highlight (i) that obedient (vs. disobedient) participants exerted an inhibitory control over their affective responses to the victim's pain, (ii) that the increase in theta power cannot be attributed to a lack of attention and involvement in the experimental procedure within one of the group.

Keywords

EEG, obedience to authority, emotion down-regulation, theta waves

Introduction

Experimental approach of obedience to authority

The Milgram's obedience experiments (Milgram, 1963, 1965, 1974) are one of the most impressive inquiries conducted in social psychology. This important piece of research is still widely cited when explaining destructive conducts such as torture (e.g., Atran, 2003; Kelman & Hamilton, 1989; Lankford, 2009; Waller, 2002). However, the Milgram's studies have also been a target of ethical criticism (e.g., Baumrind, 1964; Miller, 1986) and the evolution of ethical standards has discouraged researchers from replicating and extending Milgram's work (e.g., Elms, 1995, 2009).

In the Milgram paradigm, participant is told that he is participating in a study on the effects of punishment on learning. The "learner" (a confederate) is strapped to a chair with an electrode attached to his wrist. He has to learn a list of word pairs and whenever he makes an error, he will receive electric shocks of increasing intensity. After watching the "learner" strapped to the chair, participant is taken in an adjacent room and seated in front of a shock generator with thirty switches, ranging from 15 to 450 volts in 15-volt increments. The generator also bears verbal designations ranging from "Slight Shock" to "Danger: Severe Shock" and "XXX". Participant is instructed to administrate a shock when the "learner" provides an incorrect answer, starting to 15 volts with a 15-volt increment for each successive incorrect answer. At 75 volts, the "learner" grunts, and at 120 volts he complains verbally. At 150, the "learner" demands to be released from the experiment. These protests continue as the shocks increase, growing increasingly vehement and emotional. At 285 volts the learner's responses turn to an agonized scream. While psychiatric experts had predicted that less than 1% of participants would continue through to the most severe shock (450 volts), the results

showed that 65% of the sample of North American adult men administered the highest level of shock (Condition 5; Milgram, 1974).

Very few experimental studies has been conducted since the Milgram pioneering experiments and the mechanisms responsible for destructive obedience remain unknown (e.g., Blass, 1991; Cheetham et al., 2009). However, the first decade of the 21st century witnessed an increase of interest for destructive obedience and experimental study was reopened through the use of immersive environments (e.g., Cheetham et al. 2009; Dambrun & Vatiné, 2010; Slater et al., 2006)

Reopening the experimental study of destructive obedience

Recent studies have found immersive virtual environment technology (IVET) useful for studying social behavior and destructive obedience (e.g., Bailenson & Yee, 2005; Blascovich, Loomis, Beall, Swinth, Hoyt, & Bailenson, 2002; Cheetham et al., 2009; Dambrun & Vatiné, 2010; Freeman et al., 2005; Pertaub, Slater, & Barker, 2002; Slater et al., 2006). Dambrun and Vatiné (2010) developed the Immersive Video Milgram Obedience Experiment (IVMOE), an immersive environment very similar in content to the original Milgram paradigm except that the victim (an actor filmed and recorded) is seen on a screen. Using the IVMOE, the authors replicated several previous findings observed in real environments: (i) participants were more obedient when the victim was hidden (with similar rate of obedience than in the Milgram's voice feedback condition), (ii) right-wing authoritarianism was a significant predictor of obedience, and (iv) the more participants denied their own responsibility by blaming both the victim and the experimenter, the more they obeyed.

All in all, their results suggest that participants tended to respond to the situation as if it was real, that individuals who obeyed the most to authority within an immersive

environment have the same psychological dispositions as those who obeyed fully in Milgram's original paradigm, and that obedient participants in an immersive environment rationalized their acts in a similar manner that which has been demonstrated in a real situation.

Empathic resonance and self-regulation

In both past and recent obedience experiments, participants exhibited signs of emotional distress. However, whether the participants' response to the victim's pain was more akin to other-oriented empathic concern or to a self-oriented aversive state of personal distress is unclear (Cheetham et al., 2009). Research on empathy for pain demonstrates (i) that perception of the others' pain results, in the observer, in the activation of the neural network involved in the processing of firsthand experience of pain (e.g., Decety & Jackson, 2004), and (ii) that pain-related affective sharing can provoke either sympathy (an other-oriented motivated response) or personal distress (an aversive, self-oriented motivated response; Decety & Lamm, 2009). While sympathy initiates care-based moral reasoning and prosocial behavior (i.e., actions that are intended to help or benefit another individual; Decety & Svetlova, 2012), personal distress is associated with the desire to alleviate one's own but not the other's distress (Batson, 1991). A large range of studies showed that individuals with poor self-regulation abilities are biased to experience empathic over-arousal and personal distress in response to the others' pain (see Eisenberg, Fabes, & Spinrad, 2006, for a review)

Using a virtual version of the Milgram paradigm, Cheetham et al. (2009) conducted an fMRI study and found that the observation of the victim's pain within the obedience paradigm induced brain activations related to an aversive, self-oriented state of personal distress (i.e., activations in pre-defined brain areas involved in affective processing - e.g., right amygdala - but not in those commonly associated with emotion regulation - e.g., anterior cingulate

cortex). This result suggests that low self-regulatory control of the shared affect evoked by the emotional state of the victim could be responsible (at least partially) for destructive obedience.

Emotion down-regulation processes

Emotion regulation strategy of avoidance (e.g., inhibition of internal responses, cognitive reappraisal) is a common maladaptive response to aversive emotional experiences (e.g., Aldao, Dixon-Gordon, & De Los Reyes, 2015). Cognitive reappraisal “is a strategy used to change the emotional impact of a situation or event by reinterpreting its meaning” (Ertl, Hildebrandt, Ourina, Leicht, & Mulert, 2013). According to Goubert, Craig and Buysse (2009), individuals responding with high distress to the other’s pain could be motivated “to underestimate the observed person’s pain, in an attempt to keep their own distress within acceptable limits”. As a consequence, individuals may tend not to fully attend to the other’s experience and fail to display sympathetic responses. Brain imaging studies found a decreased amygdala (AMY) activity during the observation and the experience of pain that indicate a cognitive strategy to attenuate pain-related stress responses in a “no-escape situation” (e.g., Petrovic, Carlsson, Petersson, Hansson, & Ingvar, 2004; Simon et al., 2006). Because decreased activity within the AMY has been linked to decreased harm aversion (e.g., Blair, 2007; Moll & de Oliveira-Souza, 2007), such an inhibitory control could have detrimental consequences in a Milgram-like paradigm (i.e., increased destructive obedience).

The role of frontal theta oscillations

Research on cognitive reappraisal of negative stimuli revealed the prominent role of AMY–prefrontal cortex (PFC) interaction in emotion down-regulation (e.g., Banks et al., 2007; Kanske et al., 2011; Ochsner et al., 2004; Schulze et al., 2011). Empirical studies using fMRI showed a negative correlation between the AMY and prefrontal structures, and revealed

an inhibitory influence of the PFC on the AMY and insula during cognitive reappraisal tasks (Delgado et al., 2008; Hariri et al., 2003; Wager et al., 2008). Research in animals suggests a role of theta oscillations (3-8 Hz) as a key mechanism in the coupling between the PFC and the AMY (Lesting et al., 2011; Narayanan et al., 2011). In a recent EEG study, Ertl et al. (2013) used a set of aversive pictures to induce negative emotions in healthy participants who were asked to decrease, increase or maintain the induced emotion. The authors expected and found a significant increase in frontal theta oscillations during the regulation condition “decrease” compared to the “maintain” condition, and a positive correlation between the strength of theta power and subjective regulation success. These results suggest that frontal theta oscillations are a key physiological mechanism of emotion down-regulation by cognitive reappraisal. This is coherent with previous findings pointing out the importance of theta oscillations in the context of emotion regulation and the processing of negative stimuli (e.g., Aftanas et al., 2001; Knyazev et al., 2008).

Participants’ attention and involvement

As mentioned previously, recent studies have found IVET useful for studying destructive obedience. However, one could argue that the participants’ attention and involvement could influence their propensity to obey within an immersive environment. *Social presence* reflects the degree to which participants believe that they are “in the presence of and interacting with another veritable human being and that the behaviors of virtual humans within IVEs represent the actions of real individuals in the physical world in real time” (Blascovich et al., 2002). A lower sense of social presence could decrease the participants’ attention and involvement and increase their propensity to obey. Research demonstrates that alpha oscillations (8-12 Hz) are involved in the mechanisms of attention to environmental stimuli (e.g., Pfurtscheller & Lopes da Silva, 1999). For instance, Laufs et al.

(2003) found in a multi-modal study a negative correlation of localized BOLD activity measured via fMRI with alpha power in lateral frontal and parietal cortices that are known to support attentional processes. According to the authors, these findings suggest that alpha oscillations signal a neural baseline with “inattention”. Using H₂ ¹⁵O PET imaging, Cook et al. (1998) also found that alpha power is inversely related to underlying neural activity. Thus, if the participants’ attention and involvement in the experimental procedure could predict obedience within an immersive environment, we should find higher alpha power among obedient individuals.

In the present study, we used an alternative version of the IVMOE (Dambrun & Vatiné, 2010). Participants were instructed to administer a series of increasingly severe electric shocks to the “learner” while their electroencephalographic activity was concurrently recorded. We expected higher theta power in frontal areas as well as lower alpha power in occipito-parietal areas among obedient individuals during the obedience procedure.

Material and method

Participants

The study was approved by the local French ethics committee (ID RCB: 2014-A01823-44) and conducted according to the principles expressed in the Helsinki Declaration. 50 healthy adults (29 females, $M_{age} = 31$, $SD = 6.02$) volunteered for the experiment by signing an informed consent.

They were assessed for eligibility by applying a number of exclusion criteria: (i) familiarity with the Milgram’s research and replications; (ii) severe visual or hearing impairment, specific language impairment (e.g., aphasia, dysphasia, dysarthria, stuttering), a

history or presence of neurologic or psychiatric disease; (iii) medical treatment likely to modulate neuronal activity (e.g., benzodiazepine, antidepressant, neuroleptic, lithium).

Procedure

Recruitment and questionnaire. Participants were recruited through newspaper advertisements, public notices, flyers and direct email solicitation, to participate in a research on “memory and cardiac function”. Those interested were assessed for eligibility through an online form.

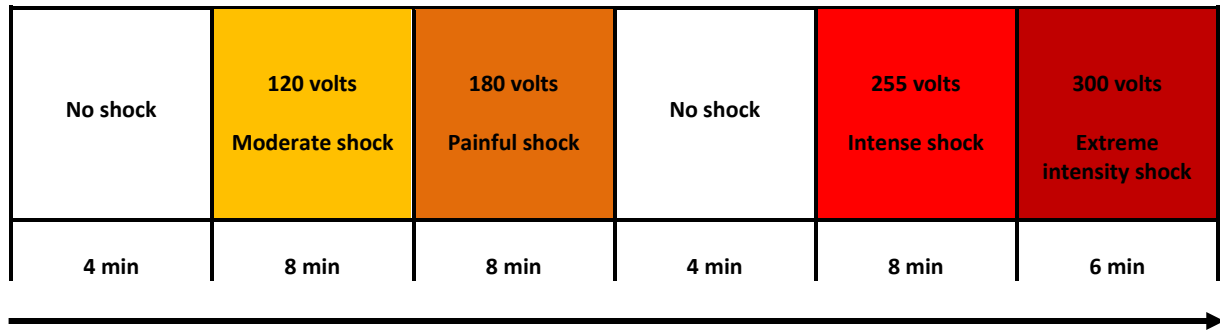
Obedience paradigm. We used an alternative version of the Immersive Video Milgram Obedience Experiment (IVMOE; Dambrun & Vatiné, 2010). First, participants were told by the experimenter (30 years-old male in a white technician’s coat) that the experiment was a simulation, as follow: “This experiment will be role-play. You will interact with another person; this is an actor that will simulate reactions to the experimental procedure. We ask you to behave exactly as if everything was real. You will have to make decisions during the experiment; we ask you to make these decisions not by considering that it will have no consequences because of the role-play, but by considering that each of your decision will have concrete consequences on real persons within a real situation. This is the most important point: you have to behave as if everything was real. If you are ok with this, right now I will behave myself as if everything was real. If you ask me about the role play, I will respond that everything is real”.

Then, participants were given a booklet about the purpose of the experiment (cover-story) and the experimental procedure. First, participants were informed that they were involved in an experiment on the effects of defibrillation on human memory; it was claimed that very little is known about the short-term effects of defibrillation on memory and how much these effects could bias medical diagnosis (i.e., misattribution of memory impairment).

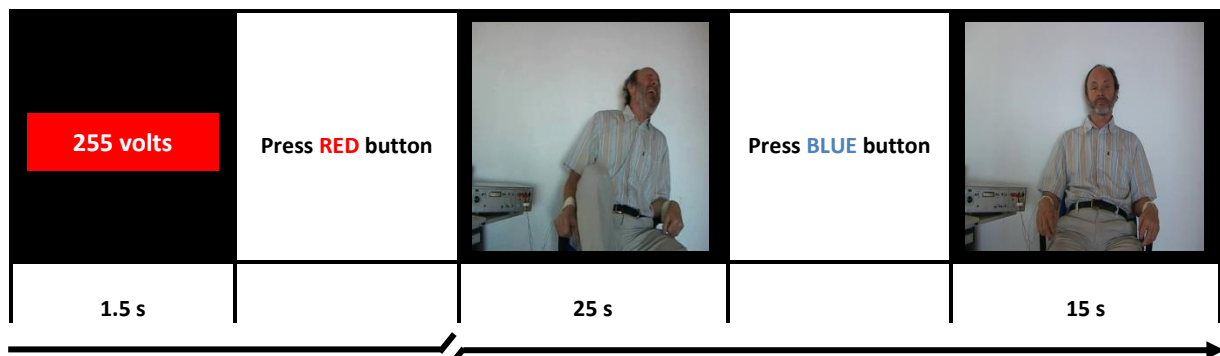
Participants were informed that we were bringing together a number of adults; we were asking some of them to pass a memory test and to receive electric shocks during this test (the “learners”), and some of them to administrate the shocks (the participants). The “learners” and the participants were supposed to be installed into two adjacent rooms. A video-conferencing system was supposed to ensure the communication. Then, participants were informed about the experimental procedure. The memory test was a paired-associate learning task. Before the experiments, the “learners” were supposed to memorize a series of word pairs; during the experiment, they were to indicate for each pair, which of four words had originally been paired with the first word. It was emphasized that to ensure the safety of the procedure (i) the memory test was administrated by a doctor, (ii) the shocks were delivered on the arms (and not on the chest), (iii) “learners” were strapped on a chair to avoid any dangerous movements. The notice provided very simple instructions to the participants: they were asked to follow the instructions appearing on the screen during the experiment (“press red button” to deliver the shock, “press blue button” to go to the next question). The last part provided Figure 1(a) with the following explanations: (i) the experimental procedure was divided into six blocks, (ii) participants were asked to administrate a shock before each question into four blocks. It took approximately five minutes to the participants to read and memorize the booklet content. After having read the booklet, the participants were invited to ask any questions about the procedure. Then, they gave their written consent for the participation. Finally, participants were instructed to avoid body movements during the experimental procedure.

The IVMOE started with the seating of the student on the “electric” chair by the virtual doctor (35 years-old male in a white technician’s coat – 1.25 min): the learner was strapped into the chair and an electrode was attached on each arm (“Installation” block). The next phase was the experimental procedure. In our immersive obedience paradigm, each trial begins by indicating the shock intensity (e.g., 120V) to be administrated to the learner (see

Figure 1(b)). At the level of 120V, the learner exhibited some cues of pain-related behavior (moan, facial expression of pain and corresponding body movement). At the level 180V, the learner's pain became more intense. At the level of 255V, the learner reacted strongly and asked to stop the experiment because the shocks were too painful. At the level of 300V, the learner's groaning was loud and asked to stop the experiment. The delay between the video onset and the shock delivery was set to vary between 3500ms to 11000ms.



(a)



(b)

Figure 1. Experimental procedure. (a) Overview of the experimental procedure. The experimental procedure was structured into six blocks, with five trials (i.e. five questions to the learner) in the first five blocks, and three trials in the last block. The total experimental time was approximately 40 min. (b) Overview of the timeline of a pain trial. Each trial begins by indicating the shock intensity (e.g., 120 volts) to be administered to the learner. In the pain trials, participants had to press the red button and the subsequent suffering of the learner was observed. Each shock is accompanied by an “electric” buzz sound. Then participants had to press the blue button. The words were enunciated by the virtual doctor and the learner gave his answer. In the non-pain trials, participants had to press the blue button and the learner was simply observed giving his answer (30 s). All parameters were computer-programmed in order to give the impression of real time and a dynamic environment.

When participants expressed their desire to stop the experiment, the experimenter intervened with four consecutive injunctions: (i) “Please continue”, (ii) “It is important for the experiment that you continue”, (iii) “The procedure requires that you continue”, (iv) “You have no other choice, you must go on”. If participants insisted on stopping the experiment after the fourth injunction, the experiment was terminated. Otherwise, this four-part sequence was restarted each time participants asked whether they could stop the experiment.

Obedience was measured as follow: (i) absolute obedience was indicated by whether the participants delivered all the shocks (0 = *not fully obedient*, 1 = *fully obedient*), (ii) the relative obedience refers to the maximum number of shocks administrated (from 0 to 18). Participants were fully debriefed after the experiment about their experience during the obedience procedure with particular reference to the impact of observing the learner in pain.

EEG recording

The acquisition of data was performed in Grenoble EEG facility IRMaGE. The electroencephalographic (EEG) activity was continuously recorded during the experimental procedure by using 64 active channels EEG system (Brain Products GmbH, Germany). The EEG caps were positioned according to the extended 10–20 system (Jasper, 1958; Oostenveld & Praamstra, 2001). The default reference and ground electrodes for this system were used during data acquisition (i.e. close to the vertex (FCz) for the reference electrode, and on the forehead (AFz) for the ground electrode). Blinks and eye movements were recorded using 4 electro-oculography electrodes placed in bipolar montages. The electrodes’ impedance was kept below 5 k Ω during all the experiment. Data were amplified, high-pass and low-pass filtered (default DC correction and anti-aliasing filters), and digitized at a sampling frequency of 1kHz.

Time-Frequency analysis

EEG signals were processed using the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and other custom scripts written in Matlab (The MathWorks Inc., USA).

Pre-processing

The EEG signal from participants' electrodes was first epoched by trials of 22s (2s before the stimulus onset to 20s after stimulus onset). Epoches were detrended and band-pass filtered (3-20 Hz). Artifacts related to ocular movements were removed by using an independent component analysis (ICA). Data were then visually inspected. All segments containing any remaining ocular activity, muscular activity or non-physiological artifacts were rejected.

Wavelets Transform

Time-frequency (TF) representations were processed using Morlet wavelets decomposition (Tallon-Baudry et al. 1996). The power for each frequency band between 3 and 20 Hz (using 1 Hz step) and each time sample from -2s to 20s (using 100ms step) was computed. We then normalized TF maps (Z-score) using the frequency power distribution of "Installation" block as baseline (see Cristofori et al., 2012; see also Knyazev et al., 2008; Leicht et al., 2010).

Statistics

Statistical analysis was performed between the TF maps of the two groups (Obedient vs. Disobedient). Significant differences in the theta (3-8 Hz) and alpha (8-12 Hz) powers were assessed using nonparametric tests within the whole analysis time window (-2s to 20s). The multiple comparison problem was solved by calculating a cluster-based test statistic, the

significance probability being inferred using the *Monte Carlo method* (Maris & Oostenveld, 2007). First, for every time-frequency bin the groups (Obedient vs. Disobedient) were compared by means of an independent samples T- statistic. All samples whose t-value was larger than the specified threshold ($\alpha=0.05$) were selected for the clustering steps. The maximum of the cluster-level statistics was chosen as the actual test statistic. The significance probability of each cluster was calculated by means of the *Monte Carlo method*, by computing this statistical test on 2000 random partitions of the data. Any cluster showing a p-value smaller than the critical alpha-level (0.05, two-sided test) was considered as significant. Finally, we concluded that the data in the two groups were significantly different, for the electrodes and time-frequency bins within the significant clusters.

Results

Rate of obedience

The distribution of break-off points is presented in Table 1. On the whole, 70% of participants obeyed fully. This is very close to the rates of obedience found by Milgram in his “baseline” condition (65%) and his “vocal feedback” condition (62.5%).

Table 1

Distribution of break-off points

Shock level	Voltage level	Number of subjects for whom this was the maximum shock
	Refuse	
	<u>NS1</u>	
	0	
	0	
	0	
	0	
	0	
	<u>S1</u>	
1	120	
2	120	
3	120	
4	120	1
5	120	1
	<u>S2</u>	
6	180	2
7	180	
8	180	1
9	180	1
10	180	2
	<u>NS2</u>	
	0	
	0	
	0	
	0	
	0	
	<u>S3</u>	
11	255	
12	255	3
13	255	
14	255	2
15	255	1
	<u>S4</u>	
16	300	
17	300	1
18	300	35
Percentage of obedience		70%
Mean maximum shock level		15.68

Note. NS1 = No shock 1; S1 = Shock 1; S2 = Shock 2; NS2 = No Shock 2; S3= Shock 3; S4 = Shock 4.

Electroencephalographic activity and Obedience

We evaluated whether theta and alpha oscillatory activities differ among obedient and disobedient participants. First, we explored differences in alpha oscillatory activity elicited by the obedience procedure between obedient and disobedient participants. The results revealed that alpha power among both obedient and disobedient participants was near from zero (i.e. did not differ from the baseline condition). Furthermore, we didn't found any significant differences in this frequency band between the two groups on the overall experience. Next, we evaluated the differences in theta oscillatory activity between obedient and disobedient participants. We found higher theta power on large clusters including mainly frontal, central and parietal electrodes among obedient individuals. More precisely, we found two clusters smaller than alpha (0.05) in **NS1** (negative¹ cluster 1, $p=0.01$, $t =0.60$ to 1.60 s; negative cluster 2, $p=0.01$, $t=5$ to 6 s), one cluster in **S1** ($p=0.04$, $t=2.60$ to 3.5 s), two clusters in **NS2** (negative cluster 1, $p=0.005$, $t=0.69$ to 2.69 s; negative cluster 2, $p=0.01$, $t =3.10$ to 3.80 s), two clusters in **S3** (negative cluster 1, $p=0.01$, $t=-0.80$ to -0.50 s; negative cluster 2, $p=0.03$, $t=17$ to 18.10 s). These differences were significant (i) after the stimulus onset in NS1 and NS2, (ii) before the shock delivery in S1 (i.e., during the shock latency period), (iii) after having pressed the red button and after the shock delivery for S3.

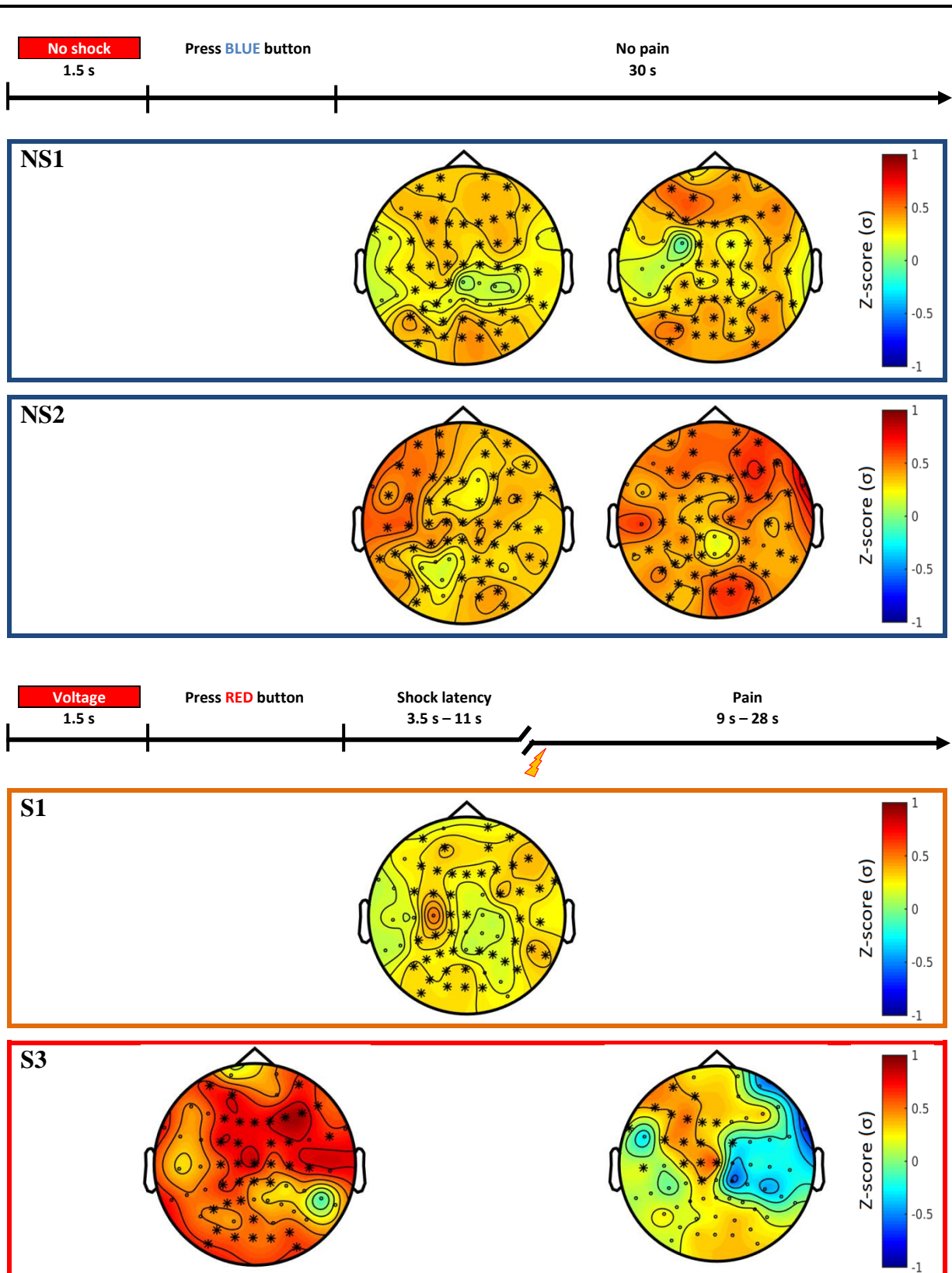


Figure 2. Contrast topographies (difference of Z scores) in the theta band power between the disobedient and obedient groups across time. Electrodes belonging to significant spatio-temporal clusters are depicted using black crosses. (1=obedient group > disobedient group; -1=obedient group < disobedient group). NS1 = No shock 1; NS2 = No Shock 2; S1 = Shock 1; S3= Shock 3.

Discussion

Based on previous research on obedience to authority (e.g., Cheetham et al., 2009) and empathy for pain (e.g., Goubert et al., 2009), we hypothesized that inhibitory control over pain sharing could be responsible for destructive obedience. Because frontal theta oscillations are believed to be key physiological mechanism of down-regulation of aversive emotional states (e.g., Ertl et al., 2013), the objective of the present study was to investigate the relation between frontal theta oscillations and destructive obedience. We predicted and found that destructive obedience was related to increased theta power. The theta activity was largely represented over the scalp surface, denoting that this increase may have come from deeper brain structures. Alpha power among both obedient and disobedient participants was near from zero, and we didn't found any significant differences between the two groups on the overall experience. A limitation is the "shock latency" jitter: the delay between the video onset and the shock delivery was not constant between all the videos. Such a latency jitter could have blurred evoked neuronal activities. Another limitation is the sample size decrease during the experiment, especially for S3 and S4 (see Table 2). The ability to detect differences between obedient and disobedient participants was undermined in these two blocks.

Our results suggest that inhibitory control precedes the onset aversive stimulus in S1 and S3, which is consistent with research showing that "some forms of emotion regulation occur proactively, for instance, when people avoid an upcoming situation that is expected to elicit an undesired emotion" (Koole, 2009). Obedient participants regulated their emotions proactively (i.e., before the emotion-arousing situation has actually occurred) but also after the shock delivery in S3. The victim's reactions to the shocks could explain this finding: while the victim exhibits some cues of pain-related behavior in S1, he reacts very strongly in S3 and asks to stop the experiment after each shock. Interestingly, a recent meta-analysis of

eight of the Milgram's obedience experiments showed that disobedience is most likely to the point at which the victim first request to be released (150 volts in the Milgram paradigm; Packer, 2008). Our results suggest that obedient participants were particularly motivated to exert an inhibitory control over their emotional responses during the periods in which the victim's protests become more vehement and emotional.

We also found higher theta power among obedient participants in NS1 and NS2. One could hypothesize that obedient participants were anticipating the pain trials during the no pain trials. Studies found that anticipating an emotional experience leads to a partial simulation of that experience (e.g., Niedenthal, 2007; Niedenthal, Mermillod, Maringer, & Hess, 2010). Therefore stress vulnerability already comes into play during the anticipation of unwanted emotions. The increased theta power may be part of a mechanism to attenuate stress responses.

The participants' electroencephalographic activity in the alpha band frequency highlights that all participants were equally attentive to the experimental procedure and that the increase in theta power cannot be attributed to a lack of attention and involvement in the experimental procedure within one of the group.

Our results are consistent with recent research on empathy and moral decision making (e.g., Blair, 2007; Moll & de Oliveira-Souza, 2007). Functional imaging studies found a decreased amygdala activity during the observation and the experience of pain (e.g., Petrovic, Ingvar, Stone-Elander, Petersson, & Hansson, 1999). It has been proposed that the altered activity in the amygdala in these studies indicate a cognitive strategy to adapt to a distressful and unavoidable painful event (e.g., Petrovic et al., 2004). These authors suggested that this decrease may be part of a mechanism to attenuate pain-related stress responses in a "no-escape situation". Because decreased activity within the amygdala has been linked to

decreased harm aversion (e.g., Blair, 2007), we predicted and found evidence that such a cognitive strategy could have detrimental consequences in a Milgram-like paradigm (i.e., increased destructive obedience).

Conclusion

Our results reveal higher theta power in frontal and parietal areas among obedient individuals, providing the first empirical evidence of a “neural signature” of destructive obedience. We also found evidence that we cannot predict obedience based on the participants’ attention and involvement in the experimental procedure. Because previous research suggests that obedient individuals tend to experience over-arousal and personal distress, our findings suggest that personal distress increases destructive obedience via a process of inhibitory control responsible for impaired emotional aversive responses to the victim’s pain and distress.

To further expand our understanding of the emotional basis of destructive obedience, the investigation of the hemodynamic responses in the Orbitofrontal cortex (OFC) could be highly informative. The OFC is closely connected to the amygdala and is sometimes conceptualized as part of the expanded limbic system (Nauta, 1979). This brain area is believed to be “an important mediator of automatic responses that guide motivated behavior (Damasio et al., 2000) and of implicit social–emotional appraisals” (e.g., Moll et al., 2002; see also Damasio et al., 2000; Milne & Grafman, 2001; Rolls et al., 1994), and its damage is responsible for a wide range of social emotional deficits, including impaired empathy and care-based moral reasoning (e.g., Koenigs et al., 2007; Namiki et al., 2008; Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). One could hypothesize that lower activity in the OFC would predict higher destructive obedience.

References

- Aftanas, L. I., & Golocheikine, S. A. (2001). Human anterior and frontal midline theta and lower alpha reflect emotionally positive states and internalized attention: high-resolution EEG investigation of meditation. *Neuroscience Letters*, *310*(1), 57-60. doi:10.1016/S0304-3940(01)02094-8
- Aldao, A., Dixon-Gordon, K. L., & De Los Reyes, A. (2015). Individual differences in physiological flexibility predict spontaneous avoidance. *Cognition and Emotion*, *30*(5), 985-998. doi:10.1080/02699931.2015.1042837
- Atran, S. (2003). Genesis of Suicide terrorism. *Science, American Association for the Advancement of Science*, *299*(5612), 1534-1539. doi:10.1126/science.1078854
- Bailenson, J. N., & Yee, N. (2005). Digital Chameleons. Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *American Psychological Society*, *16*(10), 814-819. doi:10.1111/j.1467-9280.2005.01619.x
- Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala frontal connectivity during emotion regulation. *Social Cognitive Affective Neuroscience*, *2*(4), 303-312. doi:10.1093/scan/nsm029
- Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.
- Baumrind, D. (1964). Some thoughts on the ethics of research: After reading Milgram's "Behavioral Study of Obedience". *American Psychologist*, *19*(6), 421-423. doi:10.1037/h0040128
- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*(9), 387-392. doi:10.1016/j.tics.2007.07.003

- Blascovich, J., Loomis J., Beall A. C., Swinth K. R., Hoyt C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychology Inquiry*, *13*(2), 103-124. doi:10.1207/S15327965PLI1302_01
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, *60*(3), 398-413. Retrieved from <http://www.stanleymilgram.com/pdf/understanding20%behavoir.pdf>
- Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *Neuroimage*, *49*(4), 3198-3209. doi:10.1016/j.neuroimage.2009.11.080
- Cheetham, M., Pedroni, A. F., Antley, A., Slater, M., & Jäncke, L. (2009). Virtual milgram: Empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Frontiers in Human Neuroscience*, *3*(29). doi:10.3389/neuro.09.029.2009
- Cohen, M. X., & Cavanagh, J. F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. *Frontiers in Psychology*, *2*, 30. doi:10.3389/fpsyg.2011.00030.
- Cohen, M. X., & Donner, T. H. (2013). Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. *Journal of Neurophysiology*, *110*(12), 2752-2763. doi:10.1152/jn.00479.2013
- Cohen, M. X., & Ridderinkhof, K. R. (2013). EEG Source Reconstruction Reveals Frontal-Parietal Dynamics of Spatial Conflict Processing. *Plos One*, *8*(2), e57293. doi:10.1371/journal.pone.0057293
- Cook, I. A., O'Hara, R., Uijtdehaage, S. H. J., Mandelkern, M., & Leuchter, A. F. (1999). Assessing the accuracy of topographic EEG mapping for determining local brain

- function. *Electroencephalography and Clinical Neurophysiology*, 107(6), 408-414.
doi:10.1016/S0013-4694(98)00092-3
- Cristofori, I., Moretti, L., Harquel, S., Posada, A., Deiana, G., Isnard, J., Mauguiere, F., & Sirigu, A. (2012). Theta Signal as the Neural Signature of Social Exclusion. *Cerebral Cortex*, 23(10), 2437-2447. doi:10.1093/cercor/bhs236
- Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, 41(2), 81-94. doi:0166-4328(90)90144-4
- Dambrun, M., & Vatiné, E. (2010). Reopening the study of extreme social behaviors: Obedience to authority within an immersive video environment. *European Journal of Social Psychology*, 40(5), 760-773. doi:10.1002/ejsp.646
- Davidson, R. J., Putnam, K. M., & Larson, C. L. (2000). Dysfunction in the neural circuitry of emotion regulation – a possible prelude to violence. *Science*, 289(5479), 591-594. doi:10.1126/science.289.5479.591
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, 3(2), 71-100. doi:10.1177/1534582304267187
- Decety, J., & Lamm, C. (2009). The biological basis of empathy. In J. T. Cacioppo & G. G. Bernston (Eds.), *Handbook of Neuroscience for the Behavioral Sciences*. New York: John Wiley and Sons.
- Decety, J., & Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience*, 2(1), 1-24. doi:10.1016/j.dcn.2011.05.003

- Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron*, 59(5), 829-838. doi:10.1016/j.neuron.2008.06.029
- Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Smith, M., & Maszk, P. (1996). The relations of children's dispositional empathy-related responding to their emotionality, regulation, and social functioning. *Developmental Psychology*, 32(2), 195-209. doi:10.1037/0012-1649.32.2.195
- Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg, & W. Damon (Eds.), *Handbook of child psychology. Vol. 3. Social, emotional, and personality development. 6th edition* (pp. 646-718). New York: Wiley.
- Eisenberg, N., Valiente, C., & Champion, C. (2004). Empathy-Related Responding: Moral, Social, and Socialization Correlates. In A. G. Miller (Ed.), *The Social Psychology of Good and Evil: Understanding Our Capacity for Kindness and Cruelty* (pp. 386-415). New York: Guilford Press.
- Elms, A. C. (1995). Obedience in Retrospect. *Journal of Social Issues*, 51(3), 21-31. doi:10.1111/j.1540-4560.1995.tb01332.x
- Elms, A. C. (2009). Obedience lite. *The American Psychologist*, 64(1), 32-6. doi:10.1037/a0014473.
- Ertl, M., Hildebrandt, M., Ourina, K., Leicht, G., & Mulert, C. (2013). Emotion regulation by cognitive reappraisal - the role of frontal theta oscillations. *Neuroimage*, 81, 412-421. doi:10.1016/j.neuroimage.
- Fabes, R. A., & Eisenberg, N. (1997). Regulatory control and adults' stress-related responses to daily life events. *Journal of Personality and Social Psychology*, 73(5), 1107-1117.
- Retrieved from http://www.researchgate.net/profile/Nancy_Eisenberg/publication/13864629_Regulato

ry_control_and_adults%27_stress

related_responses_to_daily_life_events/links/02e7e539be941aade6000000.pdf

- Freeman, W. J., Holmes, M. D., Burke, B. C., & Vanhatalo, S. (2003). Spatial spectra of scalp EEG and EMG from awake humans. *Clinical Neurophysiology*, *114*(6), 1053-1068. doi:10.1016/S1388-2457(03)00045-2
- Freeman, D., Garety, P. A., Bebbington, P. E., et al (2005). The psychology of persecutory ideation II: a virtual reality experimental study. *Journal of Nervous and Mental Disease*, *193*(5), 309-315.
- Funk, C. M., & Gazzaniga, M. S. (2009). The functional brain architecture of human morality. *Current Opinion in Neurobiology*, *19*(6), 678-681. doi:10.1016/j.conb.2009.09.011
- Jasper, H. H. (1958). The ten–twenty electrode system of the International Federation. *Electroencephalography Clinical Neurophysiology*, *10*, 367-380.
- Goubert, L., Craig, K. D., & Buysse, A. (2009). Perceiving others in pain: experimental and clinical evidence on the role of empathy. In W. Ickes & J. Decety (Eds.), *The Social Neuroscience of Empathy* (pp. 153-166), Cambridge (MA): The MIT Press.
- Gulbinaite, R., van Rijn, H., & Cohen, M. X. (2014). Fronto-parietal network oscillations reveal relationship between working memory capacity and cognitive control. *Frontiers in Human Neuroscience*, *8*, 761. doi:10.3389/fnhum.2014.00761
- Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, *49*(11), 3121-3129. doi:10.1016/j.neuropsychologia.2011.07.022.
- Hanslmayr, S., Pastotter, B., Bauml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience*, *20*(2), 215- 225. doi:10.1162/jocn.2008.20020

- Hariri, A. R., Goldberg, T. E., Mattay, V. S., Kolachana, B. S., Callicott, J. H., Egan, M. F., & Weinberger, D. R. (2003). Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. *The Journal of Neuroscience*, *23*(17), 6690-6694. Retrieved from <http://www.jneurosci.org/content/23/17/6690.long>
- Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of Obedience: Toward a Social Psychology of Authority and Responsibility*. New Haven: Yale University Press.
- Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, *23*(1), 4-41. doi:10.1080/02699930802619031
- Krakowski, M. (2003). Violence and serotonin: influence of impulse control, affect regulation, and social functioning. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *15*(3), 294-305. doi:10.1176/appi.neuropsych.15.3.294
- Knyazev, G. G., Bocharov, A. V., Slobodskaya, H. R., & Ryabichenko, T. I. (2008). Personality-linked biases in perception of emotional facial expressions. *Personality and Individual Differences*, *44*(5), 1093-1104. doi:10.1016/j.paid.2007.11.001
- Lankford, A. (2009). *Human Killing Machines: Systematic Indoctrination in Iran, Nazi Germany, Al Qaeda, and Abu Ghraib*. Lanham, MD: Lexington Books.
- Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., & Kleinschmidt, A. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *PNAS*, *100*(16), 11053-11058. doi:10.1073/pnas.1831638100
- Leicht, G., Herrmann, C., & Mulert, C. (2010). BOLD Response and EEG Gamma Oscillations. In C. Mulert & L. Lemieux (Eds.), *EEG – fMRI* (pp. 465-483). Berlin: Springer.

- Lesting, J., Narayanan, R. T., Kluge, C., Sangha, S., Seidenbecher, T., & Pape, H. C. (2011). Patterns of coupled theta activity in amygdala-hippocampal-prefrontal circuits during fear extinction. *Plos One*, 6(6), e21714. doi:10.1371/journal.pone.0021714
- Maris, E., & Oostenveld, R. J. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177-190. doi:10.1016/j.jneumeth.2007.03.024
- Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371-378. doi:10.1037/h0040525
- Milgram, S. (1965). Some Conditions of Obedience and Disobedience to Authority. *Human Relations*, 18(1), 57-76. doi:10.1177/001872676501800105
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. London: Tavistock Publications.
- Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science*. New York: Praeger.
- Moll, J., & Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Science*, 11(8), 319-321. doi:10.1016/j.tics.2007.06.001
- Narayanan, V., Heiming, R. S., Jansen, F., Lesting, J., Sachser, N., Pape, H-C, et al. (2011). Social Defeat: Impact on Fear Extinction and Amygdala-Prefrontal Cortical Theta Synchrony in 5-HTT Deficient Mice. *Plos One*, 6(7), e22600. doi:10.1371/journal.pone.0022600
- Niedenthal, P. M. (2007). Embodying emotion. *Science*, 316(5827), 1002-1005. doi:10.1126/science.1136930
- Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, 33(6), 417-433. doi:10.1017/S0140525X10000865

- Nigbur, R., Ivanova, G., & Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical Neurophysiology*, *122*(11), 2185-2194. doi:10.1016/j.clinph.2011.03.030.
- Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage*, *23*(2), 483-499. doi:10.1016/j.neuroimage.2004.06.030
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *156869*. doi:10.1155/2011/156869
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713-719.
- Packer, D. J. (2008). Identifying systematic disobedience in Milgram's obedience experiments: A meta-analytic review. *Perspectives on Psychological Science*, *3*(4), 301-304. doi:10.1111/j.1745-6924.2008.00080.x
- Pertaub, D. P., Slater, M., & Barker, C. (2002). An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence*, *11*(1), 68-78. doi:10.1162/105474602317343668
- Petrovic, P., Ingvar, M., Stone-Elander, S., Petersson, K. M., & Hansson, P. (1999). A PET activation study of dynamic mechanical allodynia in patients with mononeuropathy. *Pain*, *83*(3), 459-470.
- Petrovic, P., Carlsson, K., Petersson, K. M., Hansson, P., & Ingvar, M. (2004). Context-dependent deactivation of the amygdala during pain. *Journal of Cognitive Neuroscience*, *16*(7), 1289-1301. doi:10.1162/0898929041920469

- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, *110*(11), 1842-1857. doi:10.1016/S1388-2457(99)00141-8
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, *74*(2), 116-143. doi:10.1016/j.biopsycho.2006.06.009
- Schulze, L., Domes, G., Krüger, A., Berger, C., Fleischer, M., Prehn, K., Schmahl, C., Grossmann, A., Hauenstein, K., & Herpertz, S. C. (2011). Neuronal Correlates of Cognitive Reappraisal in Borderline Patients with Affective Instability. *Biological Psychiatry*, *69*(6), 564-573. doi:10.1016/j.biopsych.2010.10.025
- Segerstrom, S. C., & Ness, S. L. (2007). Heart rate variability indexes self-regulatory strength, effort, and fatigue. *Psychological Science*, *18*(3), 275-281. doi:10.1111/j.1467-9280.2007.01888.x
- Simon, D., Craig, K. D., Miltner, W. H., & Rainville, P. (2006). Brain responses to dynamic facial expressions of pain. *Pain*, *126*(1-3), 309-318. doi:10.1016/j.pain.2007.05.008
- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *Plos One*, *1*(1), e39. doi:10.1371/journal.pone.0000039
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernie, J. (1996). Stimulus Specificity of Phase-Locked and Non-Phase-Locked 40 Hz Visual Responses in Human. *The Journal of Neuroscience*, *16*(13), 4240-4249. Retrieved from <http://www.jneurosci.org/content/16/13/4240.long>
- Wager, D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K., N. (2008). Prefrontal-Subcortical Pathways Mediating Successful Emotion Regulation. *Neuron*, *59*(6), 1037-1050. doi:10.1016/j.neuron.2008.09.006

Waller, J. (2002). *Becoming Evil: How Ordinary People Commit Genocide and Mass Killing*.

Oxford: Oxford University Press.

Footnote 1

Obedient participants > disobedient participants

Exploration des signatures neurales de la soumission à l'autorité

L'étude 5 a mis en évidence une augmentation de la puissance thêta chez les participants obéissants (*vs* désobéissants) et une absence de différence entre les deux groupes dans l'amplitude des ondes alpha. Ces résultats montrent que les participants obéissants étaient tout aussi attentifs à la procédure d'obéissance que les participants désobéissants, et soutiennent l'hypothèse voulant que les participants obéissants exercent un contrôle inhibiteur sur leur résonance empathique. De plus, nos résultats montrent que l'augmentation de la puissance thêta chez les participants obéissants s'est produite juste avant le choc électrique dans les essais « chocs modérés », puis avant d'appuyer sur le bouton rouge et immédiatement après la délivrance du choc dans les essais « chocs intenses ». Ces résultats sont cohérents avec les travaux montrant que les mécanismes d'autorégulation peuvent intervenir de manière « proactive » (e.g., quand un individu anticipe l'apparition d'un stimulus aversif qui devrait provoquer une réponse émotionnelle non désirée ; e.g., Aspinwall & Taylor, 1997). L'étude 6 avait pour objectif l'exploration de la relation entre OD et activité hémodynamique au niveau du CPFDL, du CPFVM et du COF.

5.2. Etude fNIRS : Obéissance destructrice et intégration émotionnelle

Objectifs et présentation de l'étude

Si l'EEG de surface offre une excellente résolution temporelle (de l'ordre de la milliseconde), il présente toutefois une résolution spatiale faible (de quelques millimètres à un centimètre). Les outils d'imagerie cérébrale comme la spectroscopie proche infrarouge (« Functional Near-Infrared Spectroscopy » : fNIRS) permettent au contraire une localisation précise des aires cérébrales spécialement actives lors d'une tâche donnée. La fNIRS est actuellement la technique d'imagerie cérébrale la moins invasive et la plus adaptée à l'étude expérimentale du comportement social, bien qu'encore peu connue. Relativement aux autres techniques d'imagerie, la fNIRS possède des avantages importants : (i) elle ne nécessite que le port d'un casque et ne fait encourir aucun risque aux participants, (ii) elle ne requiert pas l'immobilisation des participants et est silencieuse, (iii) elle permet l'enregistrement de l'activité cérébrale assis devant un ordinateur, lors de la marche voire lors d'une activité physique plus intense comme la bicyclette. Ainsi, l'usage de la fNIRS peut s'effectuer dans des contextes relativement écologiques. En revanche, cet outil possède plusieurs limites. La mesure de l'activité neurale réalisée par la fNIRS est métabolique, donc indirecte. Plus précisément, l'imagerie optique permet de détecter les changements de concentrations en oxy-hémoglobine (oxy-Hb) et déoxy-hémoglobine (déoxy-Hb), et ainsi d'évaluer les changements d'oxygénation du sang (l'augmentation de l'activité neurale s'accompagne d'une vasodilatation locale augmentant le débit sanguin et l'oxygénation des tissus). Les taux d'oxy-Hb et de déoxy-Hb sont interprétés comme des indices de l'activité neurale locale et par extension de la contribution de la région cérébrale considérée à l'activité (cognitive, émotionnelle, motrice) du sujet. Toutefois, la fNIRS permet exclusivement de mesurer les variations de concentration en oxy-Hb et déoxy-Hb sur une profondeur de 1.5 centimètres

chez l'adulte. Et si la résolution temporelle de la fNIRS est supérieure à celle de l'imagerie par résonance magnétique, elle demeure inférieure à celle de l'EEG.

Un nombre croissant d'études montre que la fNIRS est particulièrement adaptée à l'étude du rôle des structures frontales dans le traitement de l'information émotionnelle (e.g., Brink et al., 2011 ; Doi, Nishitani, & Shinohara, 2013 ; Morinaga et al., 2007). L'objectif de l'étude 6 était d'explorer la relation entre OD et activité hémodynamique au niveau du CPFDL, du CPFVM et du COF. Nous nous attendions à ce qu'une moindre activité hémodynamique au niveau du CPFVM et du COF (hémisphère droit) prédise l'OD. Afin de tester cette hypothèse, les taux d'oxy-Hb et de déoxy-Hb au niveau du CPF des participants ($n = 46$) ont été enregistrés durant la procédure d'obéissance à l'aide de la fNIRS.

Manuscrit 6¹¹

¹¹ Manuscrit soumis pour publication auprès de la revue *Social Neuroscience*.

Activity in the Right Orbitofrontal-Ventromedial Prefrontal Cortex Predicts Obedience to Authority

Johan Lepage¹, Laurent Bègue¹, Laura Ferreri²⁻³, Aurélie Bugaiska⁴, Patrick Bard⁴, & Martial Mermillod⁵⁻⁶

¹ Laboratoire InterUniversitaire de Psychologie, Université Grenoble Alpes, Grenoble, France

² Department of Cognition, Development and Educational Psychology, University of Barcelona, Hospitalet de Llobregat, 08097, Barcelona, Spain

³ Cognition and Brain Plasticity Group, Bellvitge Biomedical Research Institute, Hospitalet de Llobregat, 08097, Barcelona, Spain

⁴ Laboratoire d'Etude de l'Apprentissage et du Développement, Université de Bourgogne, Dijon, France

⁵ Laboratoire de Psychologie et Neurocognition, CNRS UMR 5105, Université Grenoble Alpes, Grenoble, France

⁶ Institut Universitaire de France, Paris, France

Corresponding author

Johan Lepage, PhD Candidate: Laboratoire InterUniversitaire de Psychologie, UFR SHS, 1251 avenue Centrale, BP 47, 38040 Grenoble Cedex 9, France; E-mail: johan.lepage@gmail.com, Phone: (+33 4) 76 82 58 92, Fax: (+33 4) 76 82 56 65

Abstract

This study investigates the neural underpinnings of obedience to authority in a Milgram-like experiment. Harm aversion is a fundamental aspect of effective functioning in interpersonal interactions and research suggests that aversive emotional reactions to harmful acts rely on the ability to perceive, share and understand the others' affective states. The current study tests the hypothesis that destructive obedience stems from decreased harm aversion by examining the relation between obedience behavior and the reflexive engagement of the orbitofrontal-ventromedial and dorsolateral prefrontal cortices through multichannel fNIRS. The orbitofrontal-ventromedial prefrontal cortex (VMPFC) has been reliably associated with aversive emotional reactions to harmful actions, while the dorsolateral prefrontal cortex (DLPFC) plays an important role in cognitive "rational" control processes. Our results indicated that lower activity in the right VMPFC predicts higher destructive obedience. This result addresses long-standing questions about the underlying mechanisms of destructive obedience by suggesting that disobedience toward destructive orders stems, in part, from our broader capacity for empathy and harm aversion.

Keywords

Obedience to authority, orbitofrontal-ventromedial prefrontal cortex, empathy, harm aversion

Introduction

As emphasized by Crockett, Clark, Hauser and Robbins (2010), “considerations of harm and care are central to human morality [and] growing evidence supports the notion that empathic responses to the suffering of others are critical for motivating prosocial behavior”. However, research showed that aversive emotional reactions to harmful acts are highly variable between individuals and across situations (e.g., de Vignemont & Singer, 2006; Hein & Singer, 2008). One form of social interaction in which empathy and harm aversion may figure prominently is that played out in the Milgram paradigm.

Milgram’s Obedience Experiments

The Milgram paradigm consists of ordering a naïve participant to administer increasing severe electric shocks on a “learner” (a confederate) in the context of a learning experiment. In the Milgram’s studies (Milgram, 1963, 1965, 1974), participants were instructed to administer a shock each time the “learner” provided an incorrect answer in a word-memory test, starting to 15 volts with a 15-volt increment for each successive incorrect answer. At 75 volts, the “learner” grunted, and at 120 volts he complained verbally. At 150 volts, the “learner” asked to be released from the experiment. These protests continued as the shocks increased, growing increasingly vehement and emotional. At 285 volts the learner’s responses turned to an agonized scream. After 330 volts, the learner no longer reacted. While psychiatric experts had predicted that less than 1% of participants would continue through to the most severe shock (450 volts), the results showed that 65% of the sample of North American adult men administered the highest level of shock (Condition 5; Milgram, 1974). The Milgram’s obedience experiments are still widely cited when explaining destructive behaviors such as torture (e.g., Kelman, 2005; Lankford, 2009; Waller, 2002).

Studying obedience using immersive environments

The Milgram's studies have been a target of ethical criticism (particularly because of the deception involved; e.g., Baumrind, 1964; Miller, 1986) and the ethical standards subsequently established have prevented other researchers from extending Milgram's work (e.g., Elms, 2009). In such a context, the mechanisms responsible for destructive obedience (*i.e.*, obedience to orders that cause harm to another individual) remain unknown (e.g., Blass, 1991; Cheetham, Pedroni, Antley, Slater, & Jäncke, 2009). However, the first decade of the 21st century witnessed an increase of interest for destructive obedience and empirical study of obedience was reopened through the use of the immersive environments. Many studies have found immersive virtual environment technology (IVET) useful for studying social behavior (e.g., Bailenson & Yee, 2005; Blascovich, Loomis, Beall, Swinth, Hoyt, & Bailenson, 2002; Freeman et al., 2005; Pertaub, Slater, & Barker, 2002), and recent research suggests that IVET is well suited to the study of destructive obedience (e.g., Cheetham et al. 2009; Slater et al., 2006). Dambrun and Vatiné (2010) designed the Immersive Video Milgram Obedience Experiment (IVMOE), an immersive environment very similar in content to the original Milgram paradigm except that the victim (an actor filmed and recorded) is seen on a screen. When the victim was hidden, the level of obedience the authors obtained was similar to Milgram's. Replicating previous findings observed in real environments, they also found that participants were more obedient when the victim was hidden (vs. visible). Furthermore, the more participants negated their own responsibility by projecting responsibility on both the victim and the experimenter, the more they obeyed. Right-wing authoritarianism, a classic predictor of destructive obedience, predicted the level of obedience within the IVMOE.

In all obedience experiments, participants exhibited signs of emotional distress. Using a virtual version of the Milgram paradigm, Cheetham et al. (2009) conducted an fMRI study

and found some evidence that participants experienced a self-oriented aversive state of personal distress in response to the victim's pain (vs. other-oriented empathic concern for the victim well-being). According to Goubert, Craig and Buysse (2009), distressed individuals could be motivated "to underestimate the observed person's pain, in an attempt to keep their own distress within acceptable limits". Consistently, several fMRI studies found a deactivation in the amygdala (AMY) during the observation and the experience of pain in "no-escape" situations (e.g., Petrovic, Carlsson, Petersson, Hansson, & Ingvar, 2004; Simon, Craig, Miltner, & Rainville 2006). Because decreased activations within the cortical-limbic network including the AMY have been related to decreased aversion to harming others (e.g., Blair, 2007; Moll & de Oliveira-Souza, 2007), the investigation of the hemodynamic responses in the orbitofrontal-ventromedial prefrontal cortex (VMPFC) during the obedience procedure could be highly informative.

The orbitofrontal-ventromedial prefrontal cortex

The orbitofrontal-ventromedial cortex has been reported to be implicated in empathy (e.g., Decety & Jackson, 2004) and moral cognition (e.g., Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005). Closely connected to the limbic system, the VMPFC is believed to be "an important mediator of automatic responses that guide motivated behavior and of implicit social-emotional appraisals" (Moll et al., 2002; see also Damasio et al., 2000; Milne & Grafman, 2001; Rolls, Hornak, Wade, & McGrath, 1994). Its damage is responsible for a wide range of social emotional deficits, including impaired empathy and care-based moral reasoning (e.g., Moll, et al., 2005). The neuropsychological lesion literature indicates that the critical areas for impairment of empathy are located in the orbitofrontal cortex (OFC; e.g., Eslinger, 1998; Shamay-Tsoory et al., 2004). One consistent deficit that is observed is the failure of the OFC-lesioned patients to perceive, share and understand the others' emotions.

Studies provide strong evidence that the OFC underpins “affective theory of mind” (i.e., the capacity of individuals to appreciate the others’ emotions and feelings; Shamay-Tsoory et al., 2004). For instance, OFC damage have been associated with “affective theory of mind” impairments but not to “cognitive theory of mind” impairments (Shamay-Tsoory, Harari, Aharon-Peretz, & Levkovitz, 2010). Research on human morality demonstrates that VMPFC is a critical element of a cortical–limbic network that enables individuals to link emotional experience to moral evaluation and decision. Lesion studies showed that patients who sustain damage to the medial OFC may have exhibit impaired ability to automatically and rapidly process moral emotions in response to signs of moral violations (e.g., Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Young, Bechara, Tranel, Damasio, Hauser, & Damasio, 2010). Abnormal function or lesions to the OFC has been associated with a lack of emotional response to the potential harm of others (Koenigs et al., 2007; Namiki et al., 2008). Less frequently, lesions of the dorsolateral prefrontal cortex (DLPFC; typically of the right hemisphere) also lead to changes in empathy and moral behavior (e.g., Grattan, Bloomer, Archambault, & Eslinger, 1994).

Research provides evidence that the *right-sided component* of the VMPFC system may be critical. For instance, studies found that impaired “affective theory of mind” is associated with right VMPFC damage (e.g., Shamay-Tsoory, Tomer, Berger, Goldsher, & Aharon-Peretz, 2005). Lesions to the right VMPFC have also been associated to severe disturbances of social behavior, defective performance and impaired anticipatory skin conductance responses during the Gambling Task, profound abnormalities of emotional processing and acquired sociopathy (e.g., Tranel, Bechara & Denburg, 2002).

All in all, neuroimaging research suggests that the right-sided component of the VMPFC plays a critical role in anticipating the consequences of an action and encoding the

socio-emotional value of those consequences. When the neural circuits including this brain structure are inhibited, individuals often disregard the emotional consequences of their decision and behavior (e.g., Blair, 2007; Moll & de Oliveira-Souza, 2007).

Using fNIRS for measuring neural activation

Functional near-infrared spectroscopy (fNIRS) is an optical imaging technology designed to detect changes in the concentration of oxygenated hemoglobin (oxy-Hb) and deoxygenated hemoglobin (deoxy-Hb). fMRI and fNIRS both rely on the principle of neurovascular coupling to infer changes in neural activity. Also known as the hemodynamic response or Blood-Oxygenation-Level-Dependent (BOLD) response, the neurovascular coupling reflects the close temporal and regional linkage between neural activity and cerebral blood flow (CBF). In response to transient neural activity, nearby vessels dilate, increasing CBF which delivers oxygen to neural tissue. Optical imaging technology can measure blood oxygenation variations associated with transient neural activity (i.e., increased oxy-Hb and decreased deoxy-Hb). The acceptance of fNIRS as a novel technique for measuring neural activation might be partly attributable to several important methodological advantages: (i) fNIRS measurement is thought to impose a considerably less severe physical and psychological burden than that of existing neuroimaging techniques, (ii) the measurement of neural activation using near-infrared light is more robust with respect to exogenous noise in the environment and less susceptible to movements artifacts; (iii) more broadly, fNIRS is considered to be “a useful technique for measuring neural activation under less constrained and more ecologically valid settings” (Doi, Nishitani, & Shinohara, 2013; see also Ferreri, Bigand, Perrey, & Bugaiska, 2014; Tuscan et al., 2013).

A large range of functional NIRS studies on the role of the prefrontal cortex (PFC) in emotion processing treat the increased oxy-Hb concentration as a primary indicator of cortical

activation. For instance, Morinaga et al. (2007) examined oxy-Hb levels in the medial prefrontal cortex (MPFC) using fNIRS prior to and during anticipatory anxiety. Transient anxiety was induced by anticipation and a painful electric shock, and the authors found that right MPFC oxy-Hb was significantly increased during anticipation of the shock. Nishitani et al. (2011) investigated the mother-specific brain responses to infant facial emotions using fNIRS and found that discriminating infant facial emotions increased the relative oxy-Hb concentration in mothers' right (vs. left) PFC. Increased oxy-Hb has been also positively correlated with heart rate change when viewing trauma-related video clips (Matsuo et al., 2003; see also Matsuo et al., 2003; Moghimi et al., 2012; Roos et al., 2011).

A growing body of research focuses on the role of the OFC in emotional processing using functional NIRS. Thus, several studies found that oxy-Hb in the OFC increases following exposure to rewarding stimuli (e.g., Kida and Shinohara, 2013a; Minagawa-Kawai et al., 2009a). Brink et al. (2011) first introduced fNIRS to the field of neurodevelopmental research on empathy processing by investigating the neuronal correlates of empathic processing in children aged 4–8 years. The authors found that the processing of stories eliciting affective and cognitive empathy was associated with medial and bilateral OFC activation.

All in all, the current state-of-the-art in fNIRS research (i) provides evidence that oxy-Hb is sensitive to emotional stimulation (e.g., Doi et al., 2013; Nishitani et al., 2011), (ii) supports the valence-asymmetry hypothesis (e.g., Morinaga et al., 2007; Marumo et al., 2009; Tuscan et al., 2013), (iii) support the idea that the VMPFC is a critical element of a cortical–limbic network that enables individuals to perceive, share and understand the others' emotions (e.g., Brink et al., 2011).

Here, using an alternative version of the IVMOE (Dambrun & Vatiné, 2010), we examined oxy-Hb levels in the VMPFC and DLPFC regions. Because decreased activations within the cortical-limbic network including the VMPFC have been related to decreased aversion to harming others (e.g., Blair, 2007; Koenigs et al., 2007; Moll & de Oliveira-Souza, 2007; Namiki et al., 2008), we hypothesized that lower right VMPFC oxy-Hb levels would predict higher destructive obedience.

Material and Method

Participants

46 healthy adults (29 females, $M_{age} = 33.01$, $SD = 5.35$) volunteered for the experiment in exchange for 40€. All the participants were right-handed and native French-speakers and reported having normal or corrected-to-normal vision and hearing. None were taking medication known to affect the central nervous system. None have any knowledge of the Milgram's experiments and replications. Informed written consent was obtained from all participants prior to taking part in the experiment. The study was anonymous and fully complied with the Helsinki Declaration, Convention of the Council of Europe on Human Rights and Biomedicine (Grenoble ethics committee notice number 2014-07-04-23).

Experimental procedure

Obedience paradigm. We used an alternative version of the Immersive Video Milgram Obedience Experiment (IVMOE; Dambrun & Vatiné, 2010). First, the participants were told that the experiment was a simulation, as follow: "This experiment will be role-play. You will interact with another person; this is an actor that will simulate reactions to the experimental procedure. We ask you to behave exactly as if everything was real. You will have to make decisions during the experiment; we ask you to make these decisions not by considering that

it will have no consequences because of the role-play, but by considering that each of your decision will have concrete consequences on real persons within a real situation. This is the most important point: you have to behave as if everything was real. If you are ok with this, right now I will behave myself as if everything was real. If you ask me about the role play, I will respond that everything is real”.

Then, the participants were given a booklet about the purpose of the experiment (cover-story) and the experimental procedure. First, participants were informed that they were involved in an experiment on the effects of defibrillation on human memory; it was claimed that very little is known about the short-term effects of defibrillation on memory and how much these effects could bias medical diagnosis (i.e., misattribution of memory impairment). Participants were informed that we were bringing together a number of adults; we were asking some of them to pass a memory test and to receive electric shocks during this test (the “learners”), and some of them to administrate the shocks (the participants). The “learners” and the participants were supposed to be installed into two adjacent rooms. A video-conferencing system was supposed to ensure the communication. Then, participants were informed about the experimental procedure. The memory test was a paired-associate learning task. Before the experiments, the “learners” were supposed to memorize a series of word pairs; during the experiment, they were to indicate for each pair, which of four words had originally been paired with the first word. It was emphasized that to ensure the safety of the procedure (i) the memory test was administrated by a doctor, (ii) the shocks were delivered on the arms (and not on the chest), (iii) “learners” were strapped on a chair to avoid any dangerous movements. The notice provided very simple instructions to the participants: they were asked to follow the instructions appearing on the screen during the experiment (“press red button” to deliver the shock, “press blue button” to go to the next question). The last part provided Figure 1 with the following explanations: (i) the experimental procedure was divided into 3x3 blocks, (ii)

participants were asked to administrate a shock before each question into six blocks (see Figure 1). After having read the booklet, the participants were invited to ask any questions about the procedure.

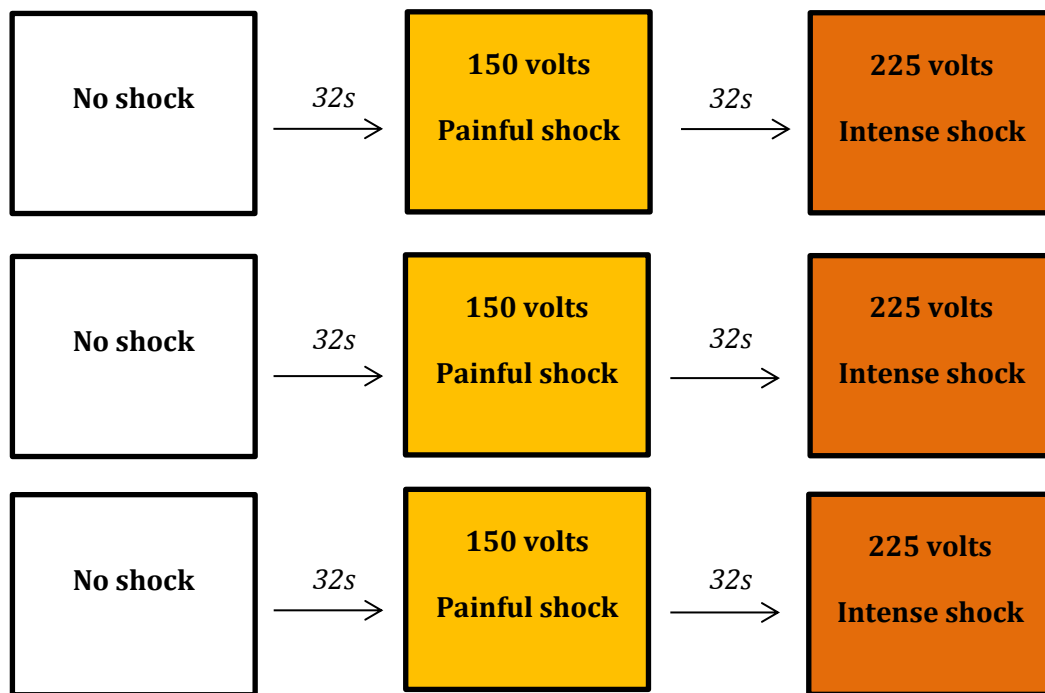


Figure 1. Overview of the experimental procedure. The experimental procedure was structured into 3x3 blocks, with 7 trials (i.e., 7 questions to the learner) in each block. The total experimental time was approximately 60 min. A rest period of 32s was inserted between each block (participants were instructed to try to relax and not to think about the task during these periods). We defined 4 conditions: “No pain” (the three “No shock” blocks), “Pain 1” (the two first “150V” and “225V” blocks), “Pain 2” (the second “150V” and “225V” blocks), “Pain 3” (the two last “150V” and “225V” blocks).

In our immersive obedience paradigm, each trial begins by indicating the shock intensity (e.g., 120V) to be administrated to the learner (see Figure 2). At the level of 150, the learner exhibited cues of pain-related behavior (moan, facial expression of pain and corresponding body movement). At the level of 225V, the learner reacted strongly and asked to stop the experiment because the shocks were too painful.

When the participants expressed their desire to stop the experiment, the experimenter intervened with four consecutive injunctions: (i) “Please continue”, (ii) “It is important for the experiment that you continue”, (iii) “The procedure requires that you continue”, (iv) “You have no other choice, you must go on”. If the participants insisted on stopping the experiment after the fourth injunction, the experiment was terminated. Otherwise, this four-part sequence was restarted each time the participants asked whether they could stop the experiment.

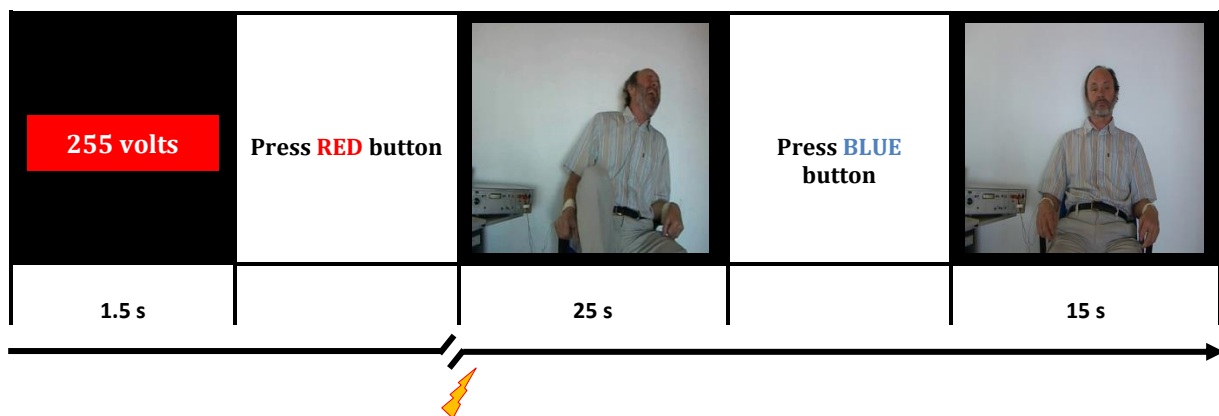


Figure 2. Overview of the timeline of a pain trial. Each trial begins by indicating the shock intensity (e.g., 120 volts) to be administrated to the learner. In the pain trials, participants had to press the red button and the subsequent suffering of the learner was observed. Each shock is accompanied by an “electric” buzz sound. Then participants had to press the blue button. The words were enunciated by the virtual doctor and the learner gave is answer. In the non-pain trials, participants had to press the blue button and the learner was simply observed giving his answer (30 s). All parameters were computer-programmed in order to give the impression of real time and a dynamic environment.

It took approximately five minutes to the participants to read and memorize the booklet content. When this step ended, participants gave their written consent for the participation. They were seated in front of a computer in a quiet, dimly lit room. After the 48 fNIRS probe-set had been fitted on the forehead overlying the PFC (see fNIRS section below for detailed description), the IVMOE started with a first rest period. Obedience was measured as follow: (i) absolute obedience was indicated by whether the participants delivered all the shocks (0 = *not fully obedient*, 1 = *fully obedient*), (ii) the relative obedience refers to the maximum number of shocks administrated (from 0 to 42). The entire experimental session,

including fNIRS recording, lasted about 75 minutes. Participants were fully debriefed after the experiment about their experience during the obedience procedure with particular reference to the impact of observing the learner in pain.

fNIRS Measurements

A 48-channel fNIRS system (OxymonMkIII, Artinis Medical Systems B.V., The Netherlands) was used to measure the concentration changes of oxy-Hb and deoxy-Hb (expressed in μM) using an age-dependent constant differential path-length factor given by $4.99 + 0.0067 * (\text{age}^{0.814})$ (Duncan et al., 1996). Data were acquired at a sampling frequency of 10 Hz. The 48 fNIRS optodes (24 emitters and 24 detectors, Figure 3) were placed symmetrically over the lateral PFC. The distance between each emitter and detector was fixed at 3 cm. For each hemisphere, fNIRS channels measured the hemoglobin concentration changes at 24 measurement points in a 12 cm² area, with the lowest optodes positioned along the Fp1-Fp2 line and the most central optodes 2 cm from the Czline (Kameyama et al., 2006), in accordance with the international 10/20 system (Okamoto et al., 2004). From top to bottom, these measurement points were labeled 1–24 (see Figure 3).

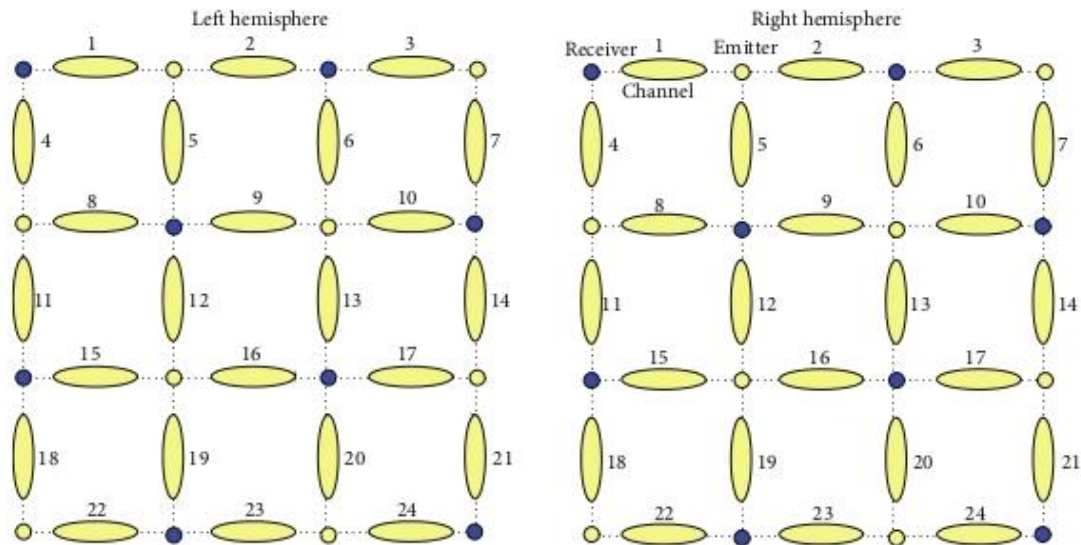


Figure 3. Channels template. The 48-channel NIRS system consisted in 8 emitters (yellow circles) and 8 receivers (blue circles) for each hemisphere, resulting in 24-left and 24-right measurement points (yellow lengthened shapes).

To optimize signal-to-noise ratio during the fNIRS recording, the 48 optodes were masked from ambient light by a black plastic cap that was kept in contact with the scalp with elastic straps, and all cables were suspended from the ceiling to minimize movement artifacts (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008). During data collection, oxy-Hb and deoxy-Hb concentration changes were displayed in real time. The signal quality and the absence of movement artifacts were verified.

Data Analysis

In order to eliminate task-irrelevant systemic physiological oscillations, the oxy-Hb and deoxy-Hb signals were first low-pass filtered (5th-order digital Butterworth filter with cut-off frequency 0.1Hz) for each of the 48 fNIRS measurement points. Data in each of the 9 blocks was baseline corrected using the mean of the oxy-Hb and deoxy-Hb signals during the last 5s of the rest phase. We sample-to-sample averaged (i.e., 10 samples/s) the baseline-

corrected signals over the blocks of each condition, yielding one average oxy-Hb and deoxy-Hb signal per participant for each condition (see Ferreri, Bigand, Bard, & Bugaiska, 2015).

Analyses were initially conducted on oxy-Hb and deoxy-Hb in a time window between 0 and 60s after stimulus onset to capture the full time course of the hemodynamic response in each block. For each participant's data, we computed four grand-averages of oxy-Hb and deoxy-Hb levels (one level of oxy-Hb and one level of deoxy-Hb per condition). According to previous studies, oxy-Hb is more useful for analyzing cortical activation than deoxy-Hb because of its higher reproducibility (Plichta et al., 2006; Sato et al., 2005), lower inter-participant variability (Sato et al., 2006) and higher correlation with the fMRI signal (Haida, 2002). Moreover, as mentioned above, research showed that oxy-Hb is more sensitive to emotional stimulation (e.g., Doi et al., 2013; Nishitani et al., 2011). Therefore, oxy-Hb was used as the main indicator of neural activation (averaging the oxy-Hb level during the whole stimulation period is a conventional analysis method; e.g., Doi et al., 2013).

First, we used the mean of oxy-Hb levels in the VMPFC (Brodmann's areas 10, 11, 12, 13, 25, 32) and the DLPFC (Brodmann's areas 9, 46) to predict destructive obedience. Based on the international 10/20 system, channels 15/16/18/19/20/22/23 were chosen for the right VMPFC, channels 16/17/19/20/21/23/24 for the left VMPFC, channels 4/5/8/9/11/12/13 for the right DLPFC, and channels 6/7/9/10/12/13/14 for the left DLPFC. We computed a mean level of oxy-Hb for each part of the VMPFC and DLPFC. The positions of the channels and clusters can be seen in Figure 4. To test our hypothesis we performed (i) a binary logistic regression analysis with the oxy-Hb levels in each part of the VMPFC and DLPFC as explanatory variables, and binary scores of obedience as the response variable, (ii) a multiple regression analysis with the oxy-Hb levels in each part of the VMPFC and DLPFC as

explanatory variables, and relative obedience as the response variable (all our continuous predictors were centered before analysis).

Second, we used the mean of oxy-Hb level in the OFC (the lateral OFC comprises Brodmann's areas 10, 11, 12, 47) to predict destructive obedience. Based on the international 10/20 system, channels 15/18/19/22 were chosen for the right OFC, and channels 17/20/21/24 were selected for the left OFC. We computed a mean level of oxy-Hb for each part of the OFC. To test our hypothesis we performed (i) a binary logistic regression analysis with the oxy-Hb levels in the right OFC and the left OFC as explanatory variables, and binary scores of obedience as the response variable, (ii) a multiple regression analysis with the oxy-Hb levels in the right OFC and the left OFC as explanatory variables, and relative obedience as the response variable (all our continuous predictors were centered before analysis).

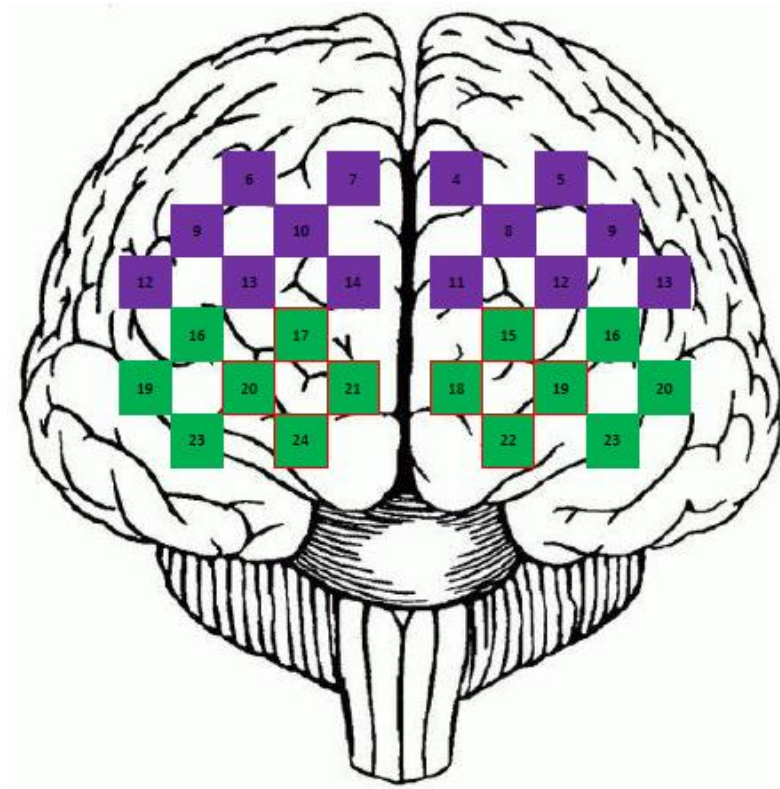


Figure 4. Channels for the orbitofrontal-ventromedial prefrontal cortex (green) and for the dorsolateral prefrontal cortex (purple). Channels for the orbitofrontal cortex are outlined in red.

Because disobedience occurred only in condition 2 (i.e., the two first “150V” and “225V” blocks), analyses were restricted to this condition and condition 1 (No Pain). Since each of the experimental conditions was presented in non-overlapping blocks, they were analyzed separately.

Results

Rate of obedience

On the whole, 76.09% of participants obeyed fully. One participant refused to administer the lowest shock and stopped the experiment. Two participants disobeyed prior to administer the first shock of 225 volts. At this level, the victim said that the shocks were too painful and asked to stop the experiment. Ten participants disobeyed during the first block of 225 volts. All the others participants obeyed fully.

Oxy-Hb levels and Obedience

Figure 5 shows the oxy-Hb levels among obedient and disobedient participants for condition 1 and condition 2.

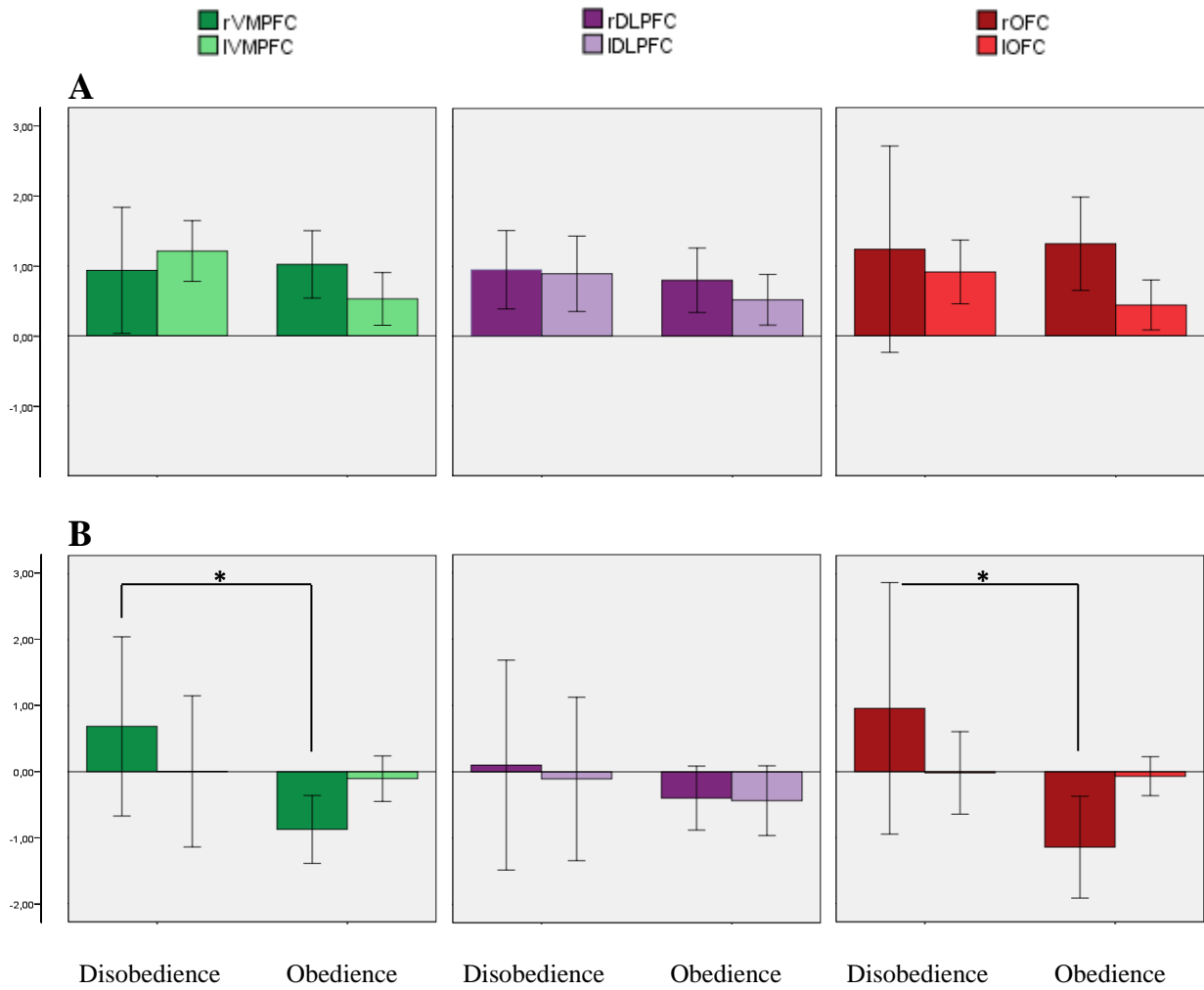


Figure 5. (A) Oxy-Hb levels among obedient and disobedient participants for condition 1 (No Pain). (B) Oxy-Hb levels among obedient and disobedient participants for condition 2 (the two first “150V” and “225V” blocks). rVMPFC = right ventromedial prefrontal cortex; IVMPFC = left ventromedial prefrontal cortex; rDLPFC = right dorsolateral prefrontal cortex; IDLPFC = left dorsolateral prefrontal cortex; rOFC = right orbitofrontal cortex; IOFC = left orbitofrontal cortex. Mean changes are shown with standard errors. (* $p < .05$, paired t -tests)

*Ventromedial and dorsolateral prefrontal cortices**Absolute obedience*

A binary logistic regression analysis performed in condition 1 showed no differences between obedient and disobedient participants. The same analysis in condition 2 revealed a significant effect of the oxy-Hb levels in the right VMPFC (higher oxy-Hb levels in the right VMPFC predict disobedience; see Table 1). There was no effect of the oxy-Hb levels in the other regions of interest.

Table 1

Binary logistic regression examining the association between oxy-Hb levels (VMPFC, DLPFC) and absolute obedience

	<i>b</i>	<i>SE</i>	<i>W</i>	<i>p</i>	95% CI
Right VMPFC	.03	1.48	5.28	.02	[-.002, .60]
Left VMPFC	2.93	1.21	.78	.37	[-.27, 31.90]
Right DLPFC	2.63	1.15	.70	.40	[-.27, 25.07]
Left DLPFC	1.03	.95	.002	.96	[-.16, 6.67]

Note. VMPFC = ventromedial prefrontal cortex; DLPFC = dorsolateral prefrontal cortex.

Relative obedience

A multiple regression analysis performed in condition 1 revealed no effects. The same analysis in condition 2 revealed that higher oxy-Hb levels in the right VMPFC predict lower relative obedience (see Table 2).

Table 2

Multiple regression examining the association between oxy-Hb levels (VMPFC, DLPFC) and relative obedience

	<i>b</i>	<i>t</i>	<i>p</i>	95% CI
Right VMPFC	-.44	-2.41	.02	[-9.25, -.81]
Left VMPFC	.16	.73	.47	[-3.35, 7.14]
Right DLPFC	.005	.02	.98	[-5.61, 5.74]
Left DLPFC	-.01	-.10	.91	[-4.75, 4.27]

Note. VMPFC = ventromedial prefrontal cortex; DLPFC = dorsolateral prefrontal cortex.

Orbitofrontal cortex

Absolute obedience

A binary logistic regression analysis performed in condition 1 showed no differences between obedient and disobedient participants. The same analysis in condition 2 revealed a significant effect of the oxy-Hb levels in the right OFC (higher oxy-Hb levels in the right OFC predict disobedience; see Table 3). There was no effect of the oxy-Hb levels in the left OFC.

Table 3

Binary logistic regression examining the association between oxy-Hb levels (right OFC, left OFC) and absolute obedience

	<i>b</i>	<i>SE</i>	<i>W</i>	<i>p</i>	95% CI
Right OFC	.14	.85	4.99	.02	[.02, .79]
Left OFC	1.40	.55	.36	.54	[-.47, 4.18]

Note. OFC = orbitofrontal cortex.

Relative obedience

A multiple regression analysis performed in condition 1 revealed no effects. The same analysis in condition 2 revealed that higher oxy-Hb levels in the right OFC predict lower relative obedience (see Table 4).

Table 4

Multiple regression examining the association between oxy-Hb levels (right OFC, left OFC) and relative obedience

	<i>b</i>	<i>t</i>	<i>p</i>	95% CI
Right OFC	-.35	-2.25	.03	[-7.54, -.39]
Left OFC	.05	.34	.73	[-2.96, 4.17]

Note. OFC = orbitofrontal cortex.

Discussion

The goal of the present study was to examine the modulatory role of emotional processing on obedience in the Milgram paradigm, a social dilemma “in which participants feel impelled to act in compliance with the experimenter’s commands to administer pain to another person while at the same time feeling driven to avoid causing harm” (Cheetham et al., 2009). Because previous research demonstrates that the VMPFC is a critical element of a cortical–limbic network that enables individuals to link emotional experience to moral evaluation and decision, we were specifically interested in determining whether VMPFC activation promotes disobedience toward destructive orders. We hypothesized and found that destructive obedience was predicted by lower activation of the right VMPFC. In line with the literature indicating that the critical areas for impairment of empathy and moral decision-making are located in the OFC, we also found that lower activation in the right OFC predicted higher destructive obedience. These findings provide preliminary evidence that insofar as the others’ well-being is concerned, the right-sided component of the VMPFC system may be critical for obedience to authority.

Brain asymmetry in emotional processing

Our results suggest that the right VMPFC has a unique role in integrating cognition and affect to produce the behavioral response to destructive orders, whereas the left-sided component may be less important. This is consistent with the models of brain asymmetry in emotional processing: the Right hemisphere hypothesis (e.g., Borod et al., 1998) and the Valence hypothesis (e.g., Davidson, 1995). The first model assumes a dominant role for the right hemisphere in emotional processing, whereas the second states that the right hemisphere is dominant for negative emotions. Studies have provided support for the Right hemisphere hypothesis. For instance, it has been found that patients with right hemisphere damage (vs.

left hemisphere damage) were more impaired in recognizing emotional facial expressions (Adolphs, Damasio, Tranel, & Damasio, 1996). Other studies found that the left hemisphere play a role in emotional processing, and the Valence model proposes that the pattern of hemispheric dominance depends on the emotional valence of the stimulus: the left hemisphere is dominant for processing positive emotions whereas the right hemisphere is dominant for processing negative emotions (e.g., Davidson, 1995). Neuroimaging research supports the Valence model. For example, Jones and Fox (1992) found a greater activation of the left hemisphere during the occurrence of positive emotion and a greater activation of the right hemisphere in the presence of negative emotion. Several fNIRS studies have also produced evidence in support of the valence asymmetry hypothesis (Morinaga et al., 2007; Marumo et al., 2009; Tuscan et al., 2013). For instance, Morinaga et al. (2007) reported that anticipation of an electrical shock was associated with a greater increase in oxy-Hb in the right (vs. left) PFC.

Cognitive control, harm aversion and destructive obedience

Current theories of destructive obedience emphasis on cognition and rationality, and downplay emotions as guides of human behavior. For instance, the agentic state model (Milgram, 1974) and the socio-cognitive theory (e.g., Bandura, 1999) emphasized on the disengagement of cognitive mechanisms responsible for moral control. Moral neuroscience research provides evidence that the right DLPFC plays an important role in cognitive “rational” control processes, while the medial prefrontal structures are responsible for more intuitive emotional reactions (e.g., Greene et al., 2004; Koenigs et al., 2007). In our study, there was no effect of the oxy-Hb levels in the DLPFC while higher activation in the right VMPFC made participants less likely to administer painful electric shocks to the victim. Because decreased activations within the cortical–limbic network including the VMPFC has

been related to decreased aversion to harming others, our results support a harm aversion account of destructive obedience.

Conclusion

Using fNIRS as a technique for measuring brain activation provides a unique approach to understanding the neural underpinnings of destructive behavior under ecologically valid settings. Our exploratory findings suggest that there is a distinct pattern of brain activity associated with destructive obedience, that higher activity in the right orbitofrontal-ventromedial prefrontal cortex promotes disobedience toward destructive orders. All in all, our results highlight the causal role of aversive emotional reactions (vs. cognitive control) on destructive obedience. This fNIRS study allows for a more rigorous investigation of the mechanisms responsible for destructive obedience.

References

- Bailenson, J. N., & Yee, N. (2005). Digital Chameleons. Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *American Psychological Society, 16*(10), 814-819. doi:10.1111/j.1467-9280.2005.01619.x
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*(3), 193-209. doi:10.1207/s15327957pspr0303_3
- Baumrind, D. (1964). Some thoughts on the ethics of research: After reading Milgram's "Behavioral Study of Obedience". *American Psychologist, 19*(6), 421-423. doi:10.1037/h0040128
- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences, 11*(9), 387-392. doi:10.1016/j.tics.2007.07.003
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychology Inquiry, 13*(2), 103-124. doi:10.1207/S15327965PLI1302_01
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology, 60*(3), 398-413. Retrieved from <http://www.stanleymilgram.com/pdf/understanding20%behavoir.pdf>
- Borod, J. C., Obler, K., L., Erhan, H., M., Grunwald, I. S., Cicero, B. A., Welkowitz, J., Santschi, C., Agisti, R. M., & Whalen, J. R. (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology, 12*(3), 446-458. doi:10.1037//0894-4105.12.3.446

- Brink, T., Urton, K., Held, D., Kirilina, E., Hofmann, M., Klann-Delius, G., Jacobs, A.M., & Kuchinke, L. (2011). The role of orbitofrontal cortex in processing empathy stories in 4-8 year-old children. *Frontiers in Psychology*, 2, 80. doi:10.3389/fpsyg.2011.00080
- Cheetham, M., Pedroni, A. F., Antley, A., Slater, M., & Jäncke, L. (2009). Virtual milgram: Empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Frontiers in Human Neuroscience*, 3(29). doi:10.3389/neuro.09.029.2009
- Ciamarelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2(2), 84-92. doi:10.1093/scan/nsm001
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, 107(40), 17433-17438. doi:10.1073/pnas.1009396107
- Damasio, A. R., Grabowski, T. J., Bechara A, Damasio, H., Ponto, L. L, Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, 3(10), 1049-1056. doi:10.1038/79871
- Dambrun, M., & Vatiné, E. (2010). Reopening the study of extreme social behaviors: Obedience to authority within an immersive video environment. *European Journal of Social Psychology*, 40(5), 760-773. doi:10.1002/ejsp.646
- Davidson, R., J. (1995). Cerebral asymmetry, emotion, and affective style. In R. J. Davidson & K. Hughdahl (Eds.), *Brain Asymmetry* (pp. 361-387). Massachusetts: MIT Press.
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, 3(2), 71-100. doi:10.1177/1534582304267187

- de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, *10*(10), 435-441. doi:10.1016/j.tics.2006.08.008
- Doi, H., Nishitani, S., & Shinohara, K. (2013). NIRS as a tool for assaying emotional function in the prefrontal cortex. *Frontiers in Human Neuroscience*, *7*, 770. doi:10.3389/fnhum.2013.00770
- Duncan, A., Meek, J. H., Clemence, M. et al. (1996). Measurement of cranial optical path length as a function of age using phase resolved near infrared spectroscopy. *Pediatric Research*, *39*(5), 889-894. doi:10.1203/00006450-199605000-00025
- Elms, A. C. (2009). Obedience lite. *The American Psychologist*, *64*(1), 32-6. doi:10.1037/a0014473.
- Eslinger, P. J. (1998). Neurological and neuropsychological bases of empathy. *European Neurology*, *39*(4), 193-199. doi:10.1159/000007933
- Ferreri, L., Bigand, E., Bard, P., & Bugaiska, A. (2015). The influence of music on prefrontal cortex during episodic encoding and retrieval of verbal information: a multichannel fNIRS study. *Behavioural Neurology*, 707625. doi:10.1155/2015/707625
- Ferreri, L., Bigand, E., Perrey, S., & Bugaiska, A. (2014). The promise of Near-Infrared Spectroscopy (NIRS) for psychological research: a brief review. *L'Année psychologique*, *114*(3), 537-569. doi:10.4074/S0003503314003054
- Freeman, D., Garety, P. A., Bebbington, P. E., Slater, M., Kuipers, E., Fowler, D., Green, C., Jordan, J., Ray, K., & Dunn, G. (2005). The psychology of persecutory ideation II: a virtual reality experimental study. *Journal of Nervous and Mental Disease*, *193*(5), 309-315.
- Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 14222-14227. doi:10.1073/pnas.0806530105

- Golkar, A., Lonsdorf, T. B., Olsson, A., Lindström, K. M., Berrebi, J., Fransson, P., Schalling, M., Ingvar, M., & Öhman, A. (2012). Distinct contributions of the dorsolateral prefrontal and orbitofrontal cortex during emotion regulation. *Plos One*, 7(11), e48107. doi:10.1371/journal.pone.0048107
- Goubert, L., Craig, K. D., & Buysse, A. (2009). Perceiving others in pain: experimental and clinical evidence on the role of empathy. In W. Ickes & J. Decety (Eds.), *The Social Neuroscience of Empathy* (pp. 153-166). Cambridge, MA: The MIT Press.
- Grattan, L. M., Bloomer, R. H., Archambault, F. X., & Eslinger, P. J. (1994). Cognitive flexibility and empathy after frontal lobe lesion. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 7, 251-257.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389-400. doi:10.1016/j.neuron.2004.09.027
- Hein, G., & Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology*, 18(2), 153-158. doi:10.1016/j.conb.2008.07.012
- Jones, N. A., & Fox, N. (1992). Electroencephalogram asymmetry during emotionally evocative films and its relation to positive and negative affectivity. *Brain Cognition*, 20(2), 280-299. doi:10.1016/0278-2626(92)90021-D
- Kameyama, M., Fukuda, M., Yamagishi, Y., Sato, T., Uehara, T., Ito, M., Suto, T., & Mikuni, M. (2006). Frontal lobe function in bipolar disorder: a multichannel near-infrared spectroscopy study. *Neuroimage*, 29(1), 172-184. doi:10.1016/j.neuroimage.2005.07.025
- Kelman, H. C. (2005). The policy context of torture: A social-psychological analysis. *International Review of the Red Cross*, 87(857), 123-134. Retrieved from

<http://scholar.harvard.edu/files/hckelman/files/PolicyContextTorture.pdf?m=13600389>

44

- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*(7138), 908-911. doi:10.1038/nature05631
- Lankford, A. (2009). *Human Killing Machines: Systematic Indoctrination in Iran, Nazi Germany, Al Qaeda, and Abu Ghraib*. Lanham, MD: Lexington Books.
- Marumo, K., Takizawa, R., Kawakubo, Y., Onitsuka, T., & Kasai, K. (2009). Gender difference in right lateral prefrontal hemodynamic response while viewing fearful faces: a multi-channel near-infrared spectroscopy study. *Neuroscience Research*, *63*(2), 89-94. doi:10.1016/j.neures.2008.10.012
- Mehta, P., & Beer, J. (2010). Neural mechanisms of the testosterone-aggression relation: The role of the orbitofrontal cortex. *Journal of Cognitive Neuroscience*, *22*(10), 2357-2368. doi:10.1162/jocn.2009.21389
- Miczek, K. A., de Almeida, R. M. M., Kravitz, E. A., Rissman, E. F., de Boer, S. F., & Raine, A. (2007). Neurobiology of Escalated Aggression and Violence. *The Journal of Neuroscience*, *27*(44), 11803-11806. doi:10.1523/JNEUROSCI.3500-07.2007
- Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology*, *67*(4), 371-378. doi:10.1037/h0040525
- Milgram, S. (1965). Some Conditions of Obedience and Disobedience to Authority. *Human Relations*, *18*(1), 57-76. doi:10.1177/001872676501800105
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. London: Tavistock Publications.
- Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science*. New York: Praeger.

- Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *The Journal of Neuroscience*, *21*(12), 1-6. Retrieved from <http://www.jneurosci.org/content/21/12/RC150.full.pdf>
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences*, *11*(8), 319-321. doi:10.1016/j.tics.2007.06.001
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, *22*(7), 2730-2736. doi:20026214
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6*(10), 799-809. doi:10.1038/nrn1768
- Morinaga, K., Akiyoshi, J., Matsushita, H., Ichioka, S., Tanaka, Y., Tsuru, J., & Hanada, H. (2007). Anticipatory anxiety-induced changes in human lateral prefrontal cortex activity. *Biological Psychology*, *74*(1), 34-38. doi:10.1016/j.biopsycho.2006.06.005
- Namiki, C., Yamada, M., Yoshida, H., Hanakawa, T., Fukuyama, H., & Murai, T. (2008). Small orbitofrontal traumatic lesions detected by high resolution MRI in a patient with major behavioural changes. *Neurocase*, *14*(6), 474-479.
- Nishitani, S., Doi, H., Koyama, A., & Shinohara, K. (2011). Differential prefrontal response to infant facial emotions in mothers compared with non-mothers. *Neuroscience Research*, *70*(2), 183-188. doi:10.1016/j.neures.2011.02.007
- Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., Oda, I., Isobe, S., Suzuki, T., Kohyama, K., & Dan, I. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10–20 system oriented for

- transcranial functional brain mapping. *Neuroimage*, 21(1), 99-111.
doi:10.1016/j.neuroimage.2003.08.026
- Pertaub, D. P., Slater, M., & Barker, C. (2002). An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence*, 11(1), 68-78.
doi:10.1162/105474602317343668
- Petrovic, P., Carlsson, K., Petersson, K. M., Hansson, P., & Ingvar, M. (2004). Context-dependent deactivation of the amygdala during pain. *Journal of Cognitive Neuroscience*, 16(7), 1289-1301. doi:10.1162/0898929041920469
- Plichta, M. M., Herrmann, M. J., Baehne, C. G., Ehlis, A-C., Richter, M. M., Pauli, P., & Fallgatter, A. J. (2006). Event-Related Functional Near-Infrared Spectroscopy (fNIRS): Are the Measurements Reliable? *Neuroimage*, 31(1), 116-124.
doi:10.1016/j.neuroimage.2005.12.008
- Rolls, E. T., Hornak, J., Wade, D., & McGrath, J. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry*, 57(12), 1518-1524.
doi:10.1136/jnnp.57.12.1518
- Sato, H., Fuchino, Y., Kiguchi, M., Katura, T., Maki, A., Yoro, T., & Koizumi, H. (2005). Intersubject Variability of Near-Infrared Spectroscopy Signals during Sensorimotor Cortex Activation. *Journal of Biomedical Optics*, 10(4), 044001.
doi:10.1117/1.1960907
- Sato, H., Kiguchi, M., Maki, A., Fuchino, Y., Obata, A., Yoro, T., & Koizumi, H. (2006). Within-Subject Reproducibility of Near-Infrared Spectroscopy Signals in Sensorimotor Activation after 6 Months. *Journal of Biomedical Optics*, 11(1), 014021. doi:10.1117/1.2166632

- Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, *46*(5), 668-677. doi:10.1016/j.cortex.2009.04.008
- Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired “affective theory of mind” is associated with right ventromedial prefrontal damage. *Cognitive Behavioral Neurology*, *18*(1), 55-67. doi:10.1097/01.wnn.0000152228.90129.99
- Shamay-Tsoory, S. G., Tomer, R., Goldsher, D., Berger, B. D., & Aharon-Peretz, J. (2004). Impairment in cognitive and affective empathy in patients with brain lesions: anatomical and cognitive correlates. *Journal of Clinical and Experimental Neuropsychology*, *26*(8), 1113-1127. doi:10.1080/13803390490515531
- Simon, D., Craig, K. D., Miltner, W. H., & Rainville, P. (2006). Brain responses to dynamic facial expressions of pain. *Pain*, *126*(1-3), 309-318. doi:10.1016/j.pain.2007.05.008
- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *Plos One*, *1*(1), e39. doi:10.1371/journal.pone.0000039
- Stalheim, E. G., Eriksson, E., Von Knorring, L., & Wide, L. (1998). Testosterone as a biological marker in psychopathy and alcoholism. *Psychiatry Research*, *77*(2), 79-88. doi:10.1016/S0165-1781(97)00143-1q
- Tranel, D., Bechara, A., & Denburg, N. L. (2002). Asymmetric functional roles of right and left ventromedial prefrontal cortices in social conduct, decision-making, and emotional processing. *Cortex*, *38*(4), 589-612. doi:10.1016/S0010-9452(08)70024-8
- Tuscan, L. A., Herbert, J. D., Forman, E. M., Juarascio, A. S., Izzetoglu, M., & Schultheis, M. (2013). Exploring frontal asymmetry using functional near-infrared spectroscopy: a

preliminary study of the effects of social anxiety during interaction and performance tasks. *Brain Imaging and Behavior*, 7(2), 140-153. doi:10.1007/s11682-012-9206-z

Virkkunen, M., Rawlings, R., Tokola, R., Poland, R. E., Guidotti, A., Nemeroff, C., Bissette, G., Kalogeras, K., Karonen, S. L., & Linnoila, M. (1994). CSF biochemistries, glucose metabolism, and diurnal activity rhythms in alcoholic, violent offenders, fire setters, and healthy volunteers. *Archives of General Psychiatry*, 51(1), 20-27.

Waller, J. (2002). *Becoming Evil: How Ordinary People Commit Genocide and Mass Killing*. Oxford: Oxford University Press.

Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, 65(6), 845-851. doi:10.1016/j.neuron.2010.03.003

L'activité du cortex préfrontal ventromédian et du cortex orbitofrontal (hémisphère droit) prédisent la soumission à l'autorité

Les résultats de l'étude 6 montrent que l'OD est associée à une moindre activité hémodynamique au sein du CPFVM et du COF (hémisphère droit), deux composantes importantes d'un réseau cortico-limbique permettant aux individus d'associer réponses émotionnelles automatiques et évaluation morale (e.g., Moll et al., 2005). Aucune relation n'a été trouvée entre OD et activité hémodynamique au niveau du CPFDL, aire cérébrale impliquée dans les processus rationnels/cognitifs contrôlés (e.g., Greene et al., 2004). Ces résultats sont cohérents avec les études montrant que l'inhibition du CPFVM et du COF droits sont à l'origine d'une large gamme de déficits socio-émotionnels (e.g., Eslinger, 1998 ; Koenigs et al., 2007 ; Shamay-Tsoory et al., 2004), notamment d'un défaut de réponse émotionnelle à l'atteinte à autrui (e.g., Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007 ; Koenigs et al., 2007; Namiki et al., 2008 ; Young, Bechara, Tranel, Damasio, Hauser, & Damasio, 2010). Ainsi, les résultats de l'étude 6 soutiennent l'idée d'un rôle causal des réactions émotionnelles aversives à l'atteinte à autrui dans le paradigme d'obéissance.

PARTIE 6. DISCUSSION GENERALE ET CONCLUSION

DISCUSSION GENERALE

6.1. Rappel du contexte général

L'étude expérimentale de l'OD initiée par Milgram (1963, 1965, 1974) a mis en évidence la forte propension des individus à obéir à des ordres menant à porter gravement atteinte à autrui, ainsi que la relation entre RWA et OD, sans pour autant apporter de réponses quant aux mécanismes responsables de l'OD. La controverse éthique provoquée par les travaux de Milgram et l'évolution des critères éthiques encadrant la recherche ont rapidement proscrit toute possibilité de réplication des expériences de Milgram. Ce contexte a motivé un important travail de réinterprétation des données existantes au cours des quatre dernières décennies (e.g., Blass, 1991 ; Burger et al., 2011 ; Haslam et al., 2014 ; Reicher & Haslam, 2011 ; Packer, 2008). Dans l'introduction du présent travail doctoral, nous avons souligné le fait que les modèles théoriques ont tous privilégié la cognition et la rationalité au détriment des émotions comme guide du comportement, un choix cohérent avec la perspective discontinuiste qui s'est développée dans la recherche sur la moralité humaine (*i.e.*, la moralité humaine serait une victoire culturelle récente sur les processus évolutifs et reposerait sur des processus cognitifs/rationnels contrôlés ; e.g., Huxley, 1894 ; Williams, 1988 ; Wright, 1994). Une autre particularité de la recherche sur l'OD est son fort ancrage dans une approche « situationniste » (*i.e.*, le comportement social serait entièrement déterminé par la situation).

Le développement de la technologie d'environnement virtuel immersif a permis de relancer l'étude expérimentale de l'OD au début des années 2000. Le développement d'approches « multiniveaux » pour la compréhension du comportement humain, comme les neurosciences sociales, a quant à lui permis de formuler des hypothèses originales sur les déterminants de l'OD. Ainsi, Cheetham et al. (2009) ont réalisé une étude IRMf portant sur la nature des réponses émotionnelles suscitées par l'observation de la victime dans une version

virtuelle du paradigme de Milgram. Les auteurs se sont appuyés sur les travaux en neurosciences sociales de l'empathie montrant (i) que la résonance empathique peut susciter de la sympathie (une réponse affective-motivationnelle orientée vers autrui) mais également de la détresse personnelle (une réponse émotionnelle aversive centrée sur soi), (ii) que la détresse personnelle résulte d'un défaut de régulation des circuits sympatho-excitateurs (e.g. Decety & Lamm, 2009 ; Eisenberg & Eggum, 2009). Cheetham et al. (2009) ont trouvé que l'observation de la douleur de la victime a provoqué des *patterns* d'activation neurale associés à la détresse personnelle (e.g., augmentation de l'activité hémodynamique de l'AMY droite). Partant de ces résultats, les auteurs ont proposé que l'OD puisse être (en partie) la conséquence d'un déficit de régulation de la détresse induite par l'observation de la douleur de la victime.

6.2. Questionnements directeurs

La question centrale de ce travail doctoral a porté sur le rôle des mécanismes d'autorégulation dans le paradigme d'obéissance. Plus précisément, nous nous sommes intéressés au rôle de la vulnérabilité au stress comme précurseur potentiel de stratégies d'autorégulation basées sur l'évitement pouvant induire une diminution de l'aversion pour l'atteinte à autrui.

Le stress est communément décrit comme un état autonome reflétant une perturbation de l'homéostasie viscérale. Il représente une réponse défensive face à une menace (réelle ou perçue) et se caractérise par une diminution de l'influence parasympathique. Des auteurs ont proposé que le stress soit un mode de fonctionnement par défaut de notre organisme assuré par les circuits sympatho-excitateurs (notamment l'AMY) orientant l'organisme vers la détection de la menace (e.g., Neuberg et al., 2011; Sandi &

Haller, 2015). Dans un environnement sûr, ce mode de fonctionnement par défaut orienté vers la menace doit être inhibé pour permettre la prosocialité.

La théorie polyvagale (e.g., Porges, 2007) met en avant le rôle de la régulation de la fonction cardiaque par le nerf vague dans la sélection des réponses face aux demandes de notre environnement. Le nerf vague myélinisé fonctionne comme un « frein vagal » permettant l'inhibition (dans un contexte perçu comme menaçant) ou la désinhibition (dans un contexte perçu comme sûr) rapide de l'influence vagale sur le cœur. La notion de « flexibilité physiologique » désigne la capacité d'un organisme à moduler rapidement l'influence parasympathique en synchronie avec les sollicitations externes. De nombreuses études empiriques ont démontré que le tonus vagal représente cette flexibilité physiologique et constitue un robuste biomarqueur de la vulnérabilité au stress. Notamment, un moindre tonus vagal a été associé à l'anxiété sociale, au stress psychosocial, à une hyperexcitation physiologique et des états émotionnels négatifs en situation de stress, et à la détresse personnelle face à la douleur d'autrui (e.g., Fabes & Eisenberg, 1997 ; Fabes et al., 1993 ; Movius & Allen, 2005 ; Thayer & Lane, 2007). La théorie polyvagale propose également qu'une surestimation chronique de la menace aura pour conséquence un défaut chronique d'inhibition des mécanismes défensifs.

Le modèle d'intégration neuroviscérale (e.g., Thayer & Lane, 2000) propose que les circuits neuronaux impliqués dans la régulation attentionnelle soient impliqués dans le contrôle de l'activité cardiaque *via* le nerf vague (Thayer et al., 2009; Park & Thayer, 2014). En l'absence de menace, le CPF exercerait une influence inhibitrice sur les circuits sympatho-excitateurs et modulerait ainsi la variabilité de la fréquence cardiaque *via* la branche myélinisée du nerf vague. De manière cohérente, la recherche empirique a mis en évidence qu'un moindre tonus vagal est associé à une hypoactivité préfrontale, une hyperactivité des

structures sous-corticale, une moindre flexibilité attentionnelle, et un biais de négativité plus important (e.g., Thayer et al., 2012). Dans cette perspective, le tonus vagal reflèterait l'influence inhibitrice du CPF sur les circuits sympatho-excitateurs. Il constituerait un biomarqueur de la perception de la menace et du stress.

Ces deux modèles proposent que l'activité vagale sous-tende la prosocialité. De manière cohérente, des études ont montré qu'un plus fort tonus vagal prédit une plus grande tendance à éprouver des sentiments prosociaux comme la sympathie et la compassion (e.g., Fabes & Eisenberg, 1997; Stellar et al., 2015). Une limite importante est le très faible nombre d'études ayant exploré la relation entre tonus vagal et mesures du comportement prosocial (voir Beffara et al., 2016). Nous avons fait l'hypothèse qu'un moindre tonus vagal devrait prédire des scores plus élevés de RWA et une plus forte propension à l'OD. Nous nous attendions également à ce que l'influence du RWA et du tonus vagal soit médiatisée par l'exercice d'un contrôle inhibiteur sur la résonance empathique. La recherche sur la régulation émotionnelle montre qu'un moindre tonus vagal prédit l'adoption spontanée de stratégies d'autorégulation basées sur l'évitement (comme l'inhibition des réponses internes) face à des stimuli aversifs (Aldao et al., 2015). Des études en imagerie cérébrale ont montré une diminution de l'activité de l'AMY durant l'observation de la douleur d'autrui dans une situation « sans issue », associée à une diminution de l'activité sympathique (e.g., Baccara et al., 2001; Petrovic et al., 1999, 2004 ; Simon et al., 2006). Ce phénomène de « désactivation » de l'AMY reflèterait une stratégie cognitive d'adaptation à des événements douloureux mais inévitables compte-tenu de la situation (Petrovic et al., 2004). Or, un moindre engagement de l'AMY dans une situation mettant en jeu le respect de l'intégrité physique d'autrui a été associé à une moindre aversion pour l'atteinte à autrui (e.g., Blair, 2007 ; Mendez, 2006). Ainsi, l'exercice d'un contrôle inhibiteur sur les réponses internes durant la procédure d'obéissance pourrait faciliter l'OD.

6.3. Principaux résultats

Autoritarisme, utilitarisme moral et tonus vagal

Les résultats de l'*étude 1* montrent que seule la dimension « dominante » du spectre autoritaire (SDO) est associée (i) à une préférence pour les options utilitaristes dans différentes versions du « dilemme du tramway », (ii) à une moindre anticipation des conséquences socio-émotionnelles des actions proposées. Les résultats de cette étude montrent également que l'influence du SDO sur la prise de décision morale est médiatisée par un moindre souci altruiste. Ainsi, le RWA ne semble pas être associé à un biais utilitariste consistant à accepter de sacrifier un individu pour le bien du plus grand nombre en raison d'une moindre capacité à éprouver des émotions prosociales (e.g., Moll et al., 2005).

Les résultats de l'*étude 2* montrent quant à eux qu'un moindre tonus vagal est associé à des scores supérieurs aux échelles RWA et SDO. Suivant la perspective polyvagale, ces résultats suggèrent qu'une surestimation chronique de la menace associée à une capacité réduite de moduler l'influence parasympathique en synchronie avec les demandes externes pourraient sous-tendre la motivation à endosser des croyances autoritaires. La perspective d'intégration neuroviscérale conduit à une conclusion similaire : une difficulté chronique à inhiber le mode de fonctionnement par défaut orienté vers la menace motiverait les individus à adopter des croyances autoritaires et des attitudes punitives.

Flexibilité physiologique et OD

Les résultats de l'*étude 3* montrent (i) que le RWA et l'OD sont associés à une moindre flexibilité physiologique (*i.e.*, moindre tonus vagal et moindre réactivité vagale), (ii) que les participants obéissants ont produit un effort cognitif (*i.e.*, augmentation du RMSSD relativement à l'état de repos) couplé à une diminution du stress physiologique (*i.e.*,

diminution de la fréquence cardiaque relativement à l'état de repos). Nous avons également trouvé que cette diminution du stress physiologique médiatise les effets de la flexibilité physiologique et du RWA sur l'OD. Ces résultats suggèrent qu'une plus grande vulnérabilité au stress est responsable de la mise en place d'un effort à l'inhibition des réponses internes, ce qui a pour effet de favoriser l'OD.

Fatigue mentale et OD

Comme pour l'étude 3, les données de l'étude 4 montre une influence du tonus vagal sur l'OD. Les résultats de cette étude montrent également que l'induction d'une fatigue mentale tend à diminuer l'OD et supprime l'influence du RWA sur l'OD. L'effet d'interaction entre RWA et fatigue mentale suggère que les personnes présentant des scores élevés de RWA sont plus obéissantes dans le paradigme de Milgram seulement si elles sont en mesure d'exercer un effort cognitif important. Dans leur ensemble, les données de l'étude 4 supportent l'hypothèse voulant que l'exercice d'un contrôle inhibiteur sur les réponses émotionnelles automatiques associées à la résonance empathique augmente l'OD. Par contraste, les réponses émotionnelles automatiques à la détresse de la victime pourraient favoriser la désobéissance à des ordres destructeurs.

Activité électroencéphalographique et OD

Les résultats de l'étude 5 mettent en évidence une augmentation de la puissance thêta chez les participants obéissants (vs désobéissants) et une absence de différence entre les deux groupes dans l'amplitude des ondes alpha. Ces résultats montrent un engagement attentionnel constant chez l'ensemble des participants et l'impossibilité de prédire leur comportement sur la base de leur implication dans la tâche. Ils soutiennent également l'hypothèse voulant que les participants obéissants exercent un effort cognitif pour diminuer leurs réponses émotionnelles aversives. Pour rappel, des études ont montré une relation positive entre

puissance thêta et diminution des ressentis d'aversion durant une tâche d'autorégulation face à des stimuli aversifs (e.g., Ertl et al., 2013). Nos résultats montrent également que cette augmentation de la puissance thêta s'est produite juste avant le choc électrique dans les essais « chocs modérés », puis avant d'appuyer sur le bouton rouge et immédiatement après la délivrance du choc dans les essais « chocs intenses ». Ce *pattern* de réponse est cohérent avec les travaux montrant que les mécanismes d'autorégulation peuvent intervenir de manière « proactive » (e.g., Aspinwall & Taylor, 1997). Ces résultats suggèrent que plus la douleur de la victime devient saillante, plus les participants obéissants ont recours au contrôle inhibiteur de leurs réponses émotionnelles.

Taux d'oxy-hémoglobine au niveau du CPF et OD

Les résultats de l'*étude 6* montrent que l'OD est associée à un moindre taux d'oxy-Hb au niveau du CPFVM (hémisphère droit), et plus particulièrement au niveau du COF (hémisphère droit). Aucune relation n'a été trouvée entre OD et taux d'oxy-Hb au niveau du CPFDL. Ces résultats sont cohérents avec les études montrant que l'inhibition du CPFVM et du COF droits sont à l'origine d'une large gamme de déficits socio-émotionnels (e.g., Koenigs et al., 2007 ; Shamay-Tsoory et al., 2004), notamment d'un défaut de réponse émotionnelle à l'atteinte à autrui (e.g., Ciaramelli et al., 2007 ; Koenigs et al., 2007; Namiki et al., 2008 ; Young et al., 2010). Cette étude soutient l'idée voulant que les réponses émotionnelles aversives à l'atteinte à autrui favorisent la désobéissance à des ordres destructeurs.

Dans leur ensemble, ces résultats soutiennent notre hypothèse de travail voulant que la vulnérabilité au stress soit (en partie) responsable de l'OD en favorisant l'exercice d'un contrôle inhibiteur sur la résonance empathique responsable d'une diminution de l'aversion pour l'atteinte à autrui.

6.4. Implications théoriques

Pris dans son ensemble, ce travail doctoral apporte une contribution théorique par l'intégration inédite d'aspects affectifs-motivationnels et biologiques à la compréhension de l'OD. Alors que la recherche sur l'OD s'ancre majoritairement dans une approche sociocognitive, nous avons ainsi privilégié une approche biopsychosociale proposant que la flexibilité physiologique (aspects biologiques) participe aux réactions à la douleur d'autrui (aspects psychologiques) qui à leur tour guident les réponses comportementales dans le paradigme de Milgram (aspects sociaux). Nos résultats soutiennent la pertinence d'une telle approche.

Par l'intégration des déterminants biologiques du comportement humain, notre travail rompt avec un courant influent des sciences sociales fondé sur l'empirisme et le constructivisme, courant s'inscrivant dans une certaine tradition philosophique affirmant qu'il n'y a pas de nature humaine et que l'humain dispose uniquement à sa naissance d'une capacité d'apprentissage générale s'exerçant dans tous les domaines (e.g., Locke, 1689). Nous souscrivons à l'idée voulant que les individus d'une même espèce soient « qualitativement semblables » : ils sont équipés d'un même ensemble de caractéristiques adaptatives définissant l'architecture typique de l'espèce humaine (e.g., Barkow, Cosmides, & Tooby, 1992). Cette architecture universelle (i) concerne autant les aspects biologiques que psychologiques, (ii) est héréditaire, et (iii) est identique du point de vue fonctionnel chez tous les individus. En revanche, les individus d'une même espèce sont « quantitativement différents » dans la mesure où ils diffèrent sur deux plans : (i) dans l'efficacité des adaptations, (ii) dans le degré auquel ils portent ces adaptations (Barkow et al., 1992). Nous souscrivons également à l'idée voulant que toute caractéristique phénotypique d'un individu

soit le produit conjoint de ses gènes et de son environnement (e.g., Hatemi & McDermott, 2011). Dans ce cadre, la nature humaine correspond à l'ensemble des adaptations physiques et psychologiques innées dont l'expression est modulée par l'environnement. Cette approche autorise la définition de modèles intégrant les déterminants exogènes *et* endogènes du comportement humain.

Il est couramment admis en sciences sociales que l'humain se distingue des autres espèces animales pour avoir transcendé son héritage évolutif (e.g., de Waal, 2003 ; Hatemi & McDermott, 2011 ; Høgh-Olesen, 2010 ; Lumsden & Wilson, 1981). Dans ce cadre, la moralité est conçue comme une victoire culturelle récente sur les processus évolutifs. Comme l'a souligné de Waal (2003), cette perspective laisse des questions majeures en suspens : pourquoi l'être humain (et seulement lui) aurait rompu avec sa biologie, comment un tel exploit est-il possible, qu'est-ce qui a motivé les humains à travers le monde à réaliser une telle rupture ? De plus, cette perspective n'est pas soutenue empiriquement. Si la moralité est une addition tardive au comportement humain, elle devrait être supportée par les aires cérébrales les plus récentes du cerveau humain et mener à des comportements spécifiques à l'espèce humaine. Or, la recherche en neurosciences montre que l'évaluation de dilemmes moraux sollicite des structures cérébrales anciennes du point de vue phylogénétique (e.g., Decety & Svetlova, 2012 ; Moll et al., 2005 ; Preston & de Waal, 2002) et la recherche comparative a mis en évidence une continuité entre primates humain et non-humain dans les sentiments et comportements moraux (e.g., sympathie, compassion, altruisme ; e.g., de Waal, 1996, 2003, 2008). L'état actuel des connaissances suggère que la morale serait la conséquence d'émotions sociales communes aux primates sociaux (e.g., Høgh-Olesen, 2010). De manière plus générale, le comportement humain reposerait moins sur des processus cognitifs/rationnels contrôlés que sur des processus émotionnels automatiques (e.g., Bargh & Chartrand, 1999 ; Zajonc, 1980, 1984).

Obstacles épistémologiques

Des auteurs ont affirmé qu'il est immoral et non-scientifique d'aborder le comportement humain sous l'angle de la biologie (e.g., Beckwith, 1993; Ledger et al. 2001). Ce rejet serait en partie la conséquence d'un amalgame entre la théorie de l'évolution (Darwin, 1859, 1871) et le Spencerisme (e.g., Spencer, 1886), un courant de pensée étendant les lois de l'évolution (principalement la « lutte pour la vie ») à l'ordre social pour en justifier les inégalités. Comme le rappelle Becquemont (2004), « ce que l'on appelle parfois encore 'Darwinisme social' est le fruit d'un malentendu : les théories désignées sous ce nom n'ont pratiquement rien à voir avec la théorie darwinienne de la sélection. Elles ont leur source dans une conception pré-darwinienne de la lutte pour la vie, que Darwin entendait dans un sens métaphorique tout autant interprétable en termes de solidarité et de dépendance. La compétition entre individus et la lutte entre races, comme moteurs de l'histoire, ne peuvent en aucune façon être qualifiées de 'Darwinisme'. » Un conflit de valeurs pourrait également expliquer le rejet d'une approche biologique du comportement humain. Comme l'a souligné Ramus (2008), « on accuse souvent la biologie d'apporter une vision déterministe de l'homme. Ce déterminisme est ressenti comme étant incompatible avec le libre arbitre, ce qui le rend insupportable à certains humanistes, qui en déduisent donc que la compréhension biologique de l'homme est nécessairement dans l'erreur. [...] il y a là une erreur patente de raisonnement. Ce n'est pas parce qu'une théorie semble incompatible avec certaines de nos valeurs qu'elle est nécessairement fausse. »

Déterminisme biologique et déterminisme social

Reconnaître que l'espèce humaine est soumise aux mêmes processus évolutifs que les autres espèces et tenter d'expliquer le comportement humain en termes d'évolution biologique ne signifie pas souscrire à un déterminisme biologique qui s'opposerait à un

déterminisme social. L'état actuel des connaissances permet d'affirmer que le comportement social est déterminé par des processus neurobiologiques complexes, notamment par les interactions SNC-SNA (e.g., Stanley & Adolphs, 2013) et que, dans le même temps, l'environnement social a un effet durable et profond sur les composantes biologiques impliquées dans la production du comportement social (e.g., Draganski, Gaser, Busch, Schuier, Bogdahn, & May, 2004 ; Maguire, Woollett, & Spiers, 2006). La recherche a notamment montré que tant les aspects fonctionnels que volumétriques du cerveau sont en partie déterminés par l'environnement cognitif et social des individus (e.g., Mermillod & Lepage, 2016).

Selon l'hypothèse du « cerveau social » (e.g., Dunbar, 1998 ; Shultz & Dunbar, 1998), la taille du cerveau des primates reflèterait la complexité des opérations cognitives exigées par la vie au sein de groupes sociaux. Le développement de certaines fonctions cognitives telles que le langage et la théorie de l'esprit semble ainsi avoir évolué sous la pression d'une vie en groupe (sous la pression à la fois du nombre des interactions et de leur complexité), ces fonctions cognitives favorisant en retour le fonctionnement du groupe *via* de nouvelles compétences à l'anticipation, la planification et la coordination. La recherche récente suggère que l'émergence des capacités d'anticipation ou de planification reposerait plus précisément, au niveau biologique, sur la mise en place d'une inter-connectivité plus riche entre le CPF et le reste du cerveau. En effet, l'augmentation du volume du CPF chez l'humain reflète une augmentation du volume de matière blanche correspondant à la myéline entourant les axones, et donc ici à une plus grande quantité de ces axones permettant la communication avec les autres régions du cerveau, comme le système limbique en charge des réponses émotionnelles. Les travaux sur le connexionnisme montrent également que, comparativement aux autres primates sociaux, le système neuronal humain bénéficie d'un faible niveau de pré-câblage lui

permettant de mieux s'adapter à un environnement changeant et imprévisible grâce à un haut niveau de plasticité synaptique (e.g., Marcus, 2009).

Des auteurs ont proposé que l'influence entre flexibilité physiologique et fonctionnement social soit circulaire (e.g., Kok & Fredrickson, 2010). De manière générale, les interactions sociales exercent une influence régulatrice sur nos réponses physiologiques et comportementales aux stimuli émotionnels (e.g., Uchino, Cacioppo, & Kiecolt-Glaser, 1996). Des travaux ont montré que l'isolement ou des interactions sociales négatives sont une puissante source de stress (e.g., House, Landis, & Umberson, 1988). La proximité sociale réduit quant à elle l'activité des circuits sympatho-exciteurs et cardio-accélérateurs (e.g., Beckes & Coan 2011 ; Brown & Brown, 2015), *via* notamment une réduction de la perception de la menace liée à l'incertitude (e.g., Feldman, 2012). De manière intéressante, les résultats de l'étude de Kok et Fredrickson (2010) suggèrent que l'activité vagale et les expériences sociales s'influencent mutuellement dans le cadre d'une boucle dynamique. Plus précisément, les auteurs ont trouvé (i) qu'un plus fort tonus vagal prédit une plus grande propension à la proximité sociale et aux émotions positives, (ii) qu'une augmentation de la proximité sociale et des émotions positives prédit une augmentation du tonus vagal.

Nous avons précisé dans notre introduction que le stress serait un mode de fonctionnement par défaut, dont l'inhibition est une précondition pour la prosocialité. Cette proposition s'appuie notamment sur la théorie polyvagale (e.g., Porges, 2007) et le modèle d'intégration neuroviscérale (e.g., Thayer & Lane, 2000). Si certaines propositions phylogénétiques de la théorie polyvagale ont été récemment remises en cause (e.g., Grossman & Taylor, 2007 ; Taylor et al., 2014), la recherche empirique soutient les principales prédictions de ces deux modèles concernant la relation entre interactions cœur-cerveau et comportement social. Les travaux les plus récents dans ce domaine ont motivé la définition de

nouveaux modèles théoriques, comme la théorie du stress comme insécurité généralisée (Brosschot et al., 2016). Cette théorie s'articule autour de six points : (i) le stress est une réponse par défaut de l'organisme, (ii) cette réponse par défaut est sous inhibition tonique du CPF tant que l'environnement est perçu comme sûr, (iii) les critères définitoires d'un environnement sûr sont appris par l'organisme au cours du temps, (iv) l'inhibition tonique du stress par le CPF module la HRV *via* la branche myélinisée du nerf vague, (v) l'influence inhibitrice du CPF consomme peu d'énergie, (vi) le stress chronique est la conséquence d'une insécurité généralisée et non de stressseurs externes. L'intérêt de ce modèle est l'unification des principales propositions théoriques existantes ainsi que l'intégration du rôle de l'apprentissage afin de mieux comprendre comment un organisme peut percevoir un environnement comme sûr. La recherche empirique soutient l'ensemble de ces propositions, des études ayant notamment montré que l'inhibition expérimentale du CPF induit une diminution du tonus vagal (Ahern et al., 2001) et que le coût métabolique de l'inhibition est moindre que celui l'excitation (Waldvogel et al., 2000). Ce dernier résultat confirme la valeur adaptative du stress comme mode de fonctionnement par défaut. Ainsi, ce modèle propose que la flexibilité physiologique réside dans l'inhibition du stress, stratégie rigide car par défaut, et propose que l'absence de menace n'implique pas que l'organisme perçoive l'environnement comme sûr. De ce fait, l'organisme peut maintenir un niveau de stress significatif en l'absence de stimuli suggérant un danger.

6.5. Limites et perspectives

Rôle causal du tonus vagal

Le fait de ne pas avoir manipulé expérimentalement le tonus vagal des participants limite nos conclusions en termes de causalité. La technique biofeedback appliquée à la HRV offre une piste pour tester le rôle causal du tonus vagal sur le RWA et l'OD. De manière

générale, le biofeedback est une technique visant à permettre aux individus d'augmenter leur capacité à réguler une fonction physiologique non perçue et non contrôlée consciemment en temps habituel à partir de signaux physiologiques enregistrés à l'aide de capteurs externes (e.g., EEG, ECG). La démarche la plus courante consiste à traiter en temps réel les signaux physiologiques de manière à les traduire en signaux visuels restitués sur un écran. Les individus peuvent alors exercer un contrôle volontaire sur leur activité physiologique en fonction des consignes fournies. Basé sur l'ECG, le biofeedback-HRV permet aux individus de suivre en temps réel l'évolution de certains paramètres cardiaques (e.g., hautes fréquences de la variabilité cardiaque ; e.g., Bornemann, Kok, Böckler, & Singer, 2016) et de les modifier en agissant directement sur eux. Le biofeedback est généralement couplé à des techniques de contrôle respiratoire (e.g., Lehrer & Gevirtz, 2014 ; Lehrer, Vaschillo, & Vaschillo, 2000). L'inspiration provoque une accélération du rythme cardiaque *via* l'inhibition de l'influence parasympathique, alors que l'expiration induit un ralentissement du rythme cardiaque *via* la stimulation du SNP (*i.e.*, l'arythmie respiratoire sinusale - RSA). Le contrôle respiratoire consiste à faire des exercices respiratoires autour d'une fréquence moyenne de 6 cycles/minutes, le but étant la synchronisation de la fréquence respiratoire et de la fréquence cardiaque (*i.e.*, cohérence cardiaque). Un protocole d'entraînement biofeedback comprend généralement 6 à 12 séances (e.g., Lehrer et al., 2000). Un certain nombre d'études montre que le biofeedback-HRV augmente les performances à des tâches cognitives, diminue la vulnérabilité au stress et favorise la prosocialité (e.g., Bornemann et al., 2016 ; Gevirtz, 2013 ; Hallman, Olsson, von Schéele, Melin, & Lyskov, 2011 ; Lehrer et al., 2004 ; Wheat & Larkin, 2010). Ainsi, nous pourrions nous attendre à ce que la participation à un protocole d'entraînement biofeedback-HRV augmente le tonus vagal des individus, diminue l'adhésion à des croyances autoritaires et favorise la désobéissance dans le paradigme d'obéissance.

Mimétisme neuromusculaire et OD

Nous avons largement fait référence dans notre travail doctoral aux mécanismes de résonance empathique et avons suggéré que le contrôle inhibiteur provoqué par la détresse personnelle diminuerait la résonance empathique. Cette résonance implique des mécanismes de mimétisme neuromusculaire, et l'électromyographie (EMG) est l'outil de mesure le plus usité pour l'étude de ce mimétisme (e.g., Neumann & Westbury, 2011). Nous n'avons pas inclus de mesure EMG à notre travail doctoral et ce point constitue autant une limite qu'une perspective stimulante.

De nombreuses études ont montré que l'observation d'expressions faciales émotionnelles (EFE) provoque chez l'observateur l'activation des muscles faciaux impliqués dans la génération des émotions exprimées par autrui (e.g., Preston & de Waal, 2002). Le terme de « contagion émotionnelle » a été utilisé pour désigner cette tendance au mimétisme neuromusculaire et la convergence émotionnelle subséquente (Hatfield, Cacioppo, & Rapson, 1994). Pour un nombre important de chercheurs, le mimétisme neuromusculaire serait une composante clé de l'empathie (e.g., de Wied, van Boxtel, Zaalberg, Goudena, & Matthys, 2006 ; Hatfield et al., 1994 ; Hoffman, 2002 ; Meltzoff & Decety, 2003 ; Preston & de Waal, 2002). Plus précisément, le mimétisme a été appréhendé comme une forme rudimentaire d'empathie et comme un facteur déterminant pour le développement des compétences de prise de perspective (e.g., Hess, Philippot, & Blairy, 1999 ; Williams et al., 2001).

Plusieurs études corrélationnelles ont trouvé une relation entre mimétisme et empathie. Par exemple, Sonnby-Borgstrom et al. (2003) ont trouvé une corrélation entre empathie dispositionnelle et mimétisme facial (les participants ayant les scores les plus élevés produisent plus de mimétisme). La recherche en endocrinologie sociale a également mis en évidence une relation entre mimétisme et empathie. Par exemple, une relation négative a été

trouvée entre niveau de testostérone (T) et empathie (e.g., Harris et al., 1996 ; Mehta & Beer, 2010), la T étant un biomarqueur de la psychopathie (Stalenheim et al., 1998) et de la personnalité antisociale (Virkkunen et al., 1994). Une étude récente a examiné la relation causale entre T et empathie *via* l'administration sublinguale d'une dose unique de T (Hermans, Putman, & van Honk, 2006). Les réponses EMG des participants à des EFE de joie et de colère ont été enregistrées et les résultats montrent que l'augmentation du niveau de T a induit une diminution du mimétisme facial (*i.e.*, atténuation du mimétisme au niveau du corrugator supercilii (CS) et du zygomatic major (ZM) ; *cf.* figure 21). Comme l'ont noté les auteurs, ces résultats sont consistants avec les modèles théoriques attribuant un rôle critique au mimétisme dans l'empathie et suggère que la T favoriserait le comportement antisocial *via* un moindre mimétisme. Enfin, la recherche clinique suggère qu'une moindre tendance au mimétisme des EFE serait un marqueur physiologique de la psychopathie (e.g., Harrison, Morgan, & Critchley, 2010). La dépression est également caractérisée par un moindre mimétisme au niveau du CS (Schwartz et al., 1978). Les réponses du CS sont aussi diminuées dans les troubles caractérisés par de faibles compétences empathiques, comme l'autisme (Hermans, van Wingen, Bos, Putman, & van Honk, 2009). D'autres études ont confirmé cette relation spécifique entre mimétisme « supérieur » (CS) et empathie dispositionnel (Harrison et al., 2010). Ainsi, il serait intéressant de mettre en relation mimétisme neuromusculaire (notamment au niveau du CS) et OD. L'enregistrement de l'activité EMG des participants devrait montrer un moindre mimétisme au niveau du CS chez les participants obéissants.

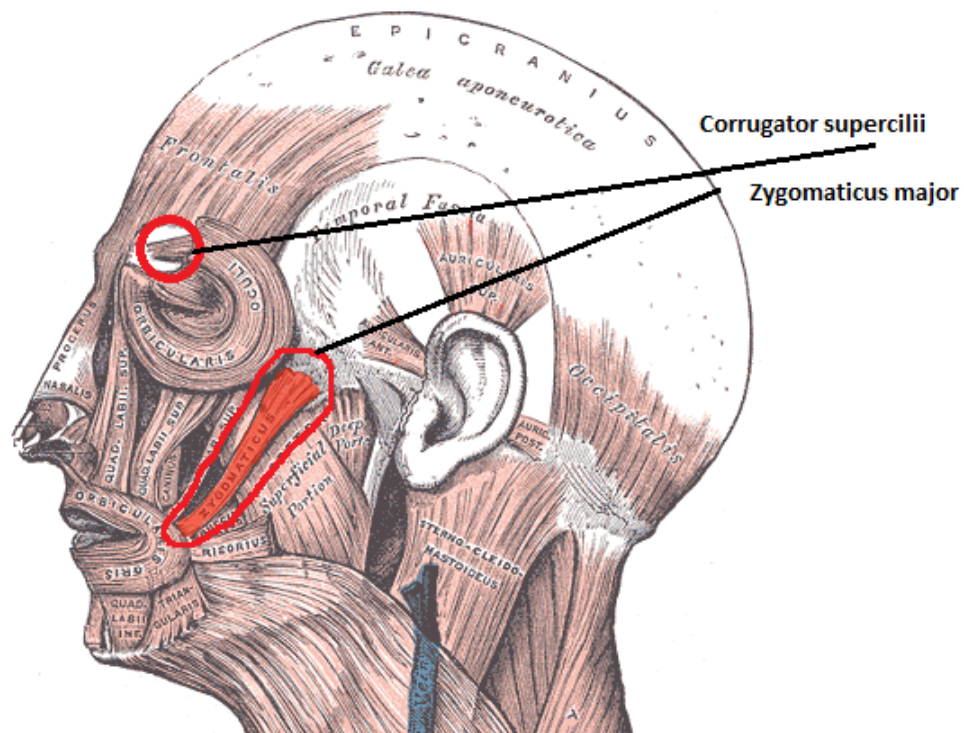


Figure 20. Le muscle zygomatique majeur (zygomaticus major) est actif face aux EFE de joie, alors que le muscle corrugateur du sourcil (corrugator supercilii) est particulièrement actif face aux EFE de colère.

La découverte des neurones miroirs dans les années 1990 (Gallese et al., 1996) fournit un mécanisme physiologique pour le lien direct entre perception et action. Cependant, certains auteurs ont récemment produit une évaluation critique de la contribution du système des neurones miroirs à l'empathie (voir Blair, 2011; Blair & Fowler, 2008, Decety, 2010, 2011). Par exemple, Blair et al. (1999) ont trouvé que si les individus imitent spontanément les EFE, il n'y aurait pas de relation entre mimétisme et reconnaissance des émotions, et le partage affectif ne semble pas être médiatisé par le mimétisme (e.g., de Vignemont & Singer, 2006). La découverte de neurones miroirs dans le cortex moteur primaire suggère que ces neurones pourraient constituer un système moteur facilitateur, agissant *via* associations apprises (e.g., Hickok, 2009). Plus récemment, des auteurs ont soutenu que la résonance motrice n'est ni suffisante ni nécessaire pour se représenter les intentions d'autrui, notamment

dans un contexte social (Decety, 2011 ; voir aussi Jacob, 2008). Actuellement, la contribution des neurones miroirs au partage affectif fait débat. Le partage affectif pourrait simplement dépendre de l'activation du « noyau affectif » qui réfère à la discrimination automatique d'un stimulus (ou de ses caractéristiques) comme appétitif ou aversif, hostile ou hospitalier, plaisant ou déplaisant, menaçant ou bienveillant (Decety, 2011 ; voir aussi Barrett, Mesquita, Ochsner, & Gross, 2007). Cet état de la littérature sur la relation entre mimétisme neuromusculaire et partage affectif explique pourquoi nous n'avons pas privilégié l'EMG dans notre travail doctoral.

Sérotonine et OD

Le comportement prosocial est communément associé à l'efficacité du contrôle volontaire des réactions émotionnelles automatiques face au stress environnemental (Davidson, Putnam, & Larson, 2000 ; Krakowski, 2003). Cependant, nous avons vu dans la partie introductive du présent travail doctoral que le comportement antisocial résulte également d'une compromission des réponses émotionnelles aversives à la détresse d'autrui (Blair, 2007 ; Damasio, Tranel, & Damasio, 1990). La recherche récente sur la sérotonine (5-HT) soutient l'idée voulant que le comportement prosocial repose sur l'aversion pour l'atteinte à autrui. La 5-HT est un neurotransmetteur présent dans le SNC. Les corps cellulaires des neurones sérotoninergiques sont localisés dans les noyaux du raphé du tronc cérébral d'où ils se projettent dans l'ensemble du cerveau et de la moelle épinière, notamment dans les aires corticales impliquées dans la morale (e.g., AMY, COF, CPFVM). Un grand nombre d'études montrent que les comportements d'affiliation et les comportements prosociaux sont associés à une fonction sérotoninergique intacte ou augmentée, alors que le comportement antisocial est associé à une compromission ou une inhibition de la fonction sérotoninergique (e.g., Crockett, 2009 ; Miczek et al., 2007). Crockett et al. (2010) ont réalisé

une étude visant à déterminer si la 5-HT favorise le comportement prosocial en améliorant la régulation émotionnelle ou au contraire en augmentant les réactions émotionnelles aversives envers l'atteinte à autrui. Les auteurs ont utilisé un inhibiteur sélectif de la recapture de la 5-HT (ISRS) et ont exposé les participants à une série de dilemmes moraux. Dans leur ensemble, les résultats montrent que l'augmentation du taux de 5-HT favorise les décisions déontologiques et le comportement prosocial en augmentant les réactions émotionnelles aversives envers l'atteinte à autrui.

Afin de tester directement l'hypothèse d'une relation entre aversion pour l'atteinte à autrui et OD, il serait intéressant d'avoir recours à un ISRS afin de manipuler le taux de 5-HT des participants en amont de la procédure d'obéissance. Si une moindre aversion pour l'atteinte à autrui favorise l'OD, l'augmentation du taux de 5-HT devrait diminuer l'OD.

CONCLUSION

Nous avons souligné dans notre introduction que les modèles théoriques mobilisés, les procédures expérimentales et la qualité des outils de mesure diffèrent selon les études sur l'OD. L'émergence et le développement d'approches comme les neurosciences sociales ouvrent la voie d'une approche « multiniveaux » pour la compréhension de l'OD dans un cadre théorique et méthodologique rigoureux et cohésif. Partant des récents travaux en neurosciences sur le jugement moral et l'empathie, nous avons fait l'hypothèse que les réponses émotionnelles automatiques associées à la résonance empathique pourraient favoriser la décision de désobéir à des ordres destructeurs. Toutefois, la recherche montre que ces réponses émotionnelles n'induisent pas de sentiments prosociaux quand elles sont associées à des niveaux d'excitation sympathique trop élevés. Partant des travaux sur l'autorégulation, nous avons fait l'hypothèse que l'OD serait la conséquence d'une moindre flexibilité physiologique responsable d'un recours à des stratégies d'autorégulation basées sur l'évitement, notamment l'inhibition des réponses internes. Les résultats que nous avons obtenus soutiennent nos hypothèses : les participants obéissants présentent une moindre flexibilité physiologique et exercent un contrôle inhibiteur sur leurs réponses internes durant la procédure d'obéissance. Il semble que ce contrôle inhibiteur soit responsable d'une moindre aversion pour l'atteinte à autrui.

A notre connaissance, notre travail de recherche est le premier à adopter une approche biopsychosociale de l'OD. L'intérêt d'une approche du comportement social par les neurosciences est régulièrement remis en cause en psychologie et plus largement en sciences sociales. Une des critiques souvent énoncées veut que l'identification des processus biologiques sous-tendant le comportement social n'apporte pas d'informations utiles à la compréhension des phénomènes sociaux. Or, hormis les aspects de localisation anatomique, la

recherche en neurosciences a permis de générer des modèles explicatifs et prédictifs d'aspects molaires du comportement social (e.g., jugement moral, empathie). Toutefois, cette démarche d'ouverture théorique implique une ouverture épistémologique, notamment l'abandon d'une certaine tradition de recherche en sciences sociales (empiriste et constructiviste) dont les principales affirmations ne sont plus compatibles avec l'état actuel des connaissances.

Une approche multiniveaux permettrait également de définir des standards méthodologiques exigeants et partagés. Relativement aux outils de mesure, les techniques de mesure physiologique permettent des mesures objectives, continues et en temps réel de processus physiologiques largement automatiques et/ou incontrôlables associés à des construits motivationnels, affectifs et cognitifs – des construits difficiles à quantifier sans biais ou artefacts en l'absence de telles mesures et impossibles à quantifier autrement qu'*in vivo* (e.g., Blascovich, 2014). En ce qui concerne les procédures expérimentales, nous avons vu dans notre introduction que la technologie d'environnement virtuel est un outil à fort potentiel pour la recherche sur le comportement humain (e.g., Bailenson & Yee, 2005 ; Blascovich et al., 2002 ; Dambrun & Vatiné, 2010 ; Freeman et al., 2003, 2005 ; Schilbach et al., 2006 ; Slater et al., 2006). D'aucun pourrait affirmer que les variations dans la « présence » pourraient sous-tendre les différences interindividuelles que nous avons observées dans l'OD (*i.e.*, une moindre « présence » pourrait diminuer l'attention et l'implication des participants et ainsi augmenter leur propension à obéir). Cependant, les résultats de notre étude EEG (étude 5) montrent un engagement attentionnel constant chez l'ensemble des participants et l'impossibilité de prédire leur comportement sur la base de leur implication dans la procédure d'obéissance.

La recherche expérimentale sur l'OD étant un domaine d'investigation peu investi par la communauté scientifique, ce travail doctoral s'est inscrit dans une démarche attentive aux

récents développements théoriques et méthodologiques tant en psychologie expérimentale qu'en neurosciences. Par notre approche biopsychosociale de l'OD et notre recours systématique à un environnement vidéo immersif, nous souhaitons contribuer à la définition d'un nouveau cadre théorique, épistémologique et méthodologique pour la recherche sur l'OD. Cette approche biopsychosociale a été amorcée par Milgram lui-même, quand il affirmait (i) que les traits psychologiques produisant notre capacité à vivre dans des groupes hiérarchiques sont la conséquence de l'histoire évolutive de l'espèce humaine, (ii) que l'OD est sous-tendue par « des changements dans les schémas du fonctionnement neural » (Milgram, 1974). Nous espérons ainsi stimuler autant que faire se peut ce domaine de recherche confronté à d'importants obstacles depuis les travaux pionniers de Milgram il y a plus d'un demi-siècle.

BIBLIOGRAPHIE

- Adams, R. B., Jr., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science*, *300*(5625), 1536. doi:10.1126/science.1082244
- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., & Sanford, R. N. (1950). *The authoritarian personality*. New York: Harper.
- Aftanas, L. I., & Golocheikine, S. A. (2001). Human anterior and frontal midline theta and lower alpha reflect emotionally positive states and internalized attention: high-resolution EEG investigation of meditation. *Neuroscience Letters*, *310*, 57-60. doi:10.1016/S0304-3940(01)02094-8
- Ahern, G. L., Sollers, J. J., Lane, R. D., Labiner, D. M., Herring, A. M., Weinand, M. E., Hutzler, R., & Thayer, J. F. (2001). Heart Rate and Heart Rate Variability Changes in the Intracarotid Sodium Amobarbital Test. *Epilepsia*, *42*, 912-921. doi:10.1046/j.1528-1157.2001.042007912.x
- Akinola, M., & Mendes, W. B. (2014). It's good to be the king: Neurobiological benefits of higher social standing. *Social Psychological and Personality Science*, *5*(1), 43-51. doi:10.1177/1948550613485604
- Akitsuki, Y., & Decety, J. (2009). Social context and perceived agency affects empathy for pain: an event-related fMRI investigation. *Neuroimage*, *47*(2), 722-734. doi:10.1016/j.neuroimage.2009.04.091
- Albrecht, M. (1973). *Conformity and dissent in the absence of consensus* (Master's thesis), University of Saint Louis, Bruxelles.
- Aldao, A., Dixon-Gordon, K. L., & De Los Reyes, A. (2015). Individual differences in physiological flexibility predict spontaneous avoidance. *Cognition and Emotion*, *30*(5), 985-998. doi:10.1080/02699931.2015.1042837
- Allen, D. F. (1983). Follow-up analysis of use of forewarning and deception in psychological experiments. *Psychological Reports*, *52*, 899-906.
- Allik, J., & McCrae, R. R. (2004). Towards a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology*, *35*, 13-28. doi:10.1177/0022022103260382
- Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Perseus Books.

- Allport, G. W., & Ross, J. M. (1967). Personal religious orientation and prejudice. *Journal of Personality and Social Psychology*, 5, 432-443. doi:10.1037/h0021212
- Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Canada: University of Manitoba Press.
- Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism*. San Francisco: Jossey-Bass.
- Altemeyer, B. (1996). *The Authoritarian Specter*. Cambridge: Harvard University Press.
- Amodio, D. M., Jost, J. T., Master, S. L., & Yee, C. M. (2007). Neurocognitive correlates of liberalism and conservatism. *Nature Neuroscience*, 10(10), 1246-1247. doi:10.1038/nn1979
- Amodio, D. M., & Ratner, K. G. (2013). The neuroscience of social cognition. In D. Carlston (Ed.) *Handbook of social cognition* (pp. 702-728). New York: Oxford University Press.
- Ancona, L., & Pareyson, R. (1968). Contributo allo studio della aggressione: La dinamica della obbedienza distruttiva. *Archivio di Psicologia. Neurologia. e Psichiatria*, 29, 340-372.
- Anderson, K. J. (1994). Impulsivity, caffeine, and task difficulty: A within-subjects test of the Yerkes-Dodson law. *Personality and Individual Differences*, 16(6), 813-829. doi:10.1016/0191-8869(94)90226-7
- Arendt, H. (1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Penguin Books.
- Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. New York: Humanities Press.
- Aronson, E., & Carlsmith, J. M. (1969). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology*. Reading, MA: Addison-Wesley.
- Aronson, E., Wilson, T. D., & Akert, R. M. (2005). *Social Psychology* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Aspinwall, L. G., & Taylor, S. E. (1997). A stitch in time: Self-regulation and proactive coping. *Psychological Bulletin*, 121, 417-436. doi:10.1037/0033-2909.121.3.417
- Atran, S. (2003). Genesis of Suicide terrorism. *Science. American Association for the Advancement of Science*, 299, 1534-1539. doi:10.1126/science.1078854
- Bailenson, J. N., & Yee, N. (2005). Digital Chameleons. Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *American Psychological Society*, 16(10), 814-819. doi:10.1111/j.1467-9280.2005.01619.x

- Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala frontal connectivity during emotion regulation. *Social Cognitive Affective Neuroscience*, 2(4), 303-312. doi:10.1093/scan/nsm029
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45-103). Hillsdale, NJ: Lawrence Erlbaum.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193-209. doi:10.1207/s15327957pspr0303_3
- Bargh, J. A., & Chartrand, T. L. (1999). The Unbearable Automaticity of Being. *American Psychologist*, 54(7), 462-479. doi:10.1037/0003-066X.54.7.462
- Barkow, J. H., Cosmides, L., & Tooby, J. (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York, NY: Oxford University Press.
- Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic (2nd ed.)*. New York: Guilford Press.
- Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, 58, 373-403. doi:10.1146/annurev.psych.58.110405.085709
- Bartels, D. M., & Pizarro, D. A. (2011). The Mismeasure of Morals: Antisocial Personality Traits Predict Utilitarian Responses to Moral Dilemmas. *Cognition*, 121, 154-161. doi:10.1016/j.cognition.2011.05.010
- Batson, C. D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.
- Baumann, N., Kaschel, R., & Kuhl, J. (2007). Affect sensitivity and affect regulation in dealing with positive and negative affect. *Journal of Research in Personality*, 41, 239-248. doi:10.1016/j.jrp.2006.05.002
- Baumeister, R. E., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology*, 74(5), 1252-1265. doi:10.1037/0022-3514.74.5.1252
- Baumeister, R. F., Sparks, E. A., Stillman, T. F., & Vohs, K. D. (2008). Free will in consumer behavior: Self-control, ego depletion, and choice. *Journal of Consumer Psychology*, 18(1), 4-13. doi:10.1016/j.jcps.2007.10.002

- Baumrind, D. (1964). Some thoughts on the ethics of research: After reading Milgram's "Behavioral Study of Obedience". *American Psychologist*, *19*, 421-423. doi:10.1037/h0040128
- Beauvois, J. L., Courbet, D., & Oberlé, D. (2012). The prescriptive power of the television host: A transposition of Milgram's obedience paradigm to the context of TV game show. *European Review of Applied Psychology*, *62*(4), 111-119. doi:10.1016/j.erap.2012.02.001
- Becerra, L., Breiter, H. C., Wise, R., Gonzalez, R. G., & Borsook, D. (2001) Reward Circuitry Activation by Noxious Thermal Stimuli. *Neuron*, *32*(5), 927-946. doi:10.1016/S0896-6273(01)00533-5
- Bechara, A. (2004). The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, *55*, 30-40. doi:10.1016/j.bandc.2003.04.001
- Beckes, L., & Coan, J. A. (2011). Social baseline theory: The role of social proximity in emotion and economy of action. *Social and Personality Psychology Compass*, *5*(12), 976-988. doi:10.1111/j.1751-9004.2011.00400.x
- Beckwith, J. (1993). A historical view of social responsibility in genetics. *BioScience*, *43*(5), 327-333. doi:10.2307/1312065
- Becquemont, D. (2004). Une régression épistémologique : le "darwinisme social". *Espaces Temps*, *84*(1), 91-105. doi:10.3406/espac.2004.4242
- Beffara, B., Bret, A.G., Vermeulen, N., & Mermillod, M. (2016). Resting High Frequency Heart Rate Variability Selectively Predicts Cooperative Behavior. *Physiology & Behavior*, *164*, 417-428. doi:10.1016/j.physbeh.2016.06.011
- Bègue, L., Beauvois, J. L., Courbet, D., Oberle, D., Lepage, J., & Duke, A. A. (2015). Personality Predicts Obedience in a Milgram Paradigm. *Journal of Personality*, *83*(3), 299-306. doi:10.1111/jopy.12104
- Bercovitch, F. B. (1991). Social stratification, social strategies, and reproductive success in primates. *Ethology and Sociobiology*, *12*, 315-333. doi:10.1016/0162-3095(91)90023-J
- Berkowitz, J. (1999). Evil is more than banal: Situationism and the concept of evil. *Personality and Social Psychology Review*, *3*, 246-253. doi:10.1207/s15327957pspr0303_7

- Bierbrauer, G. (1979). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology*, 9(1), 67-84. doi:10.1002/ejsp.2420090106
- Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, 57(1), 1-29. doi:10.1016/0010-0277(95)00676-P
- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, 11(9), 387-392. doi:10.1016/j.tics.2007.07.003
- Blair, R. J. R. (2011). Should affective arousal be grounded in perception-action coupling? *Emotion Review*, 3(1), 109-110. doi:10.1177/1754073910384157
- Blair, C., & Ursache, A. (2011). A bidirectional model of executive functions and self-regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation* (2nd ed., pp. 300-320). New York, NY: Guilford Press.
- Blair, R. J. R., & Fowler, K. (2008). Moral emotions and moral reasoning from the perspective of affective cognitive neuroscience: A selective review. *European Journal of Developmental Science*, 2, 303-323. doi:10.3233/DEV-2008-2308
- Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of Nonverbal Behavior*, 23(1), 5-41. doi:10.1023/A:1021370825283
- Blascovich, J. (2014). Using Physiological Indexes in Social Psychological Research. In H. T., Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd Ed., pp. 101-122). New York: Cambridge University Press.
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychology Inquiry*, 13(2), 103-124. doi:10.1207/S15327965PLI1302_01
- Blascovich, J., & Mendes, W. B. (2010). Social Psychophysiology and Embodiment. *Handbook of Social Psychology* (5th ed.). New York: Wiley.
- Blascovich, J., Mendes, W. B., Hunter, S. B., & Lickel, B. (2000). Stigma, threat, and social interaction. In T. F. Heatherton, M. R. Hebl, & J. G. Hull (Eds.), *The social psychology of stigma* (pp. 307-333). New York: Guilford.
- Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social "facilitation" as challenge and threat. *Journal of Personality and Social Psychology*, 77, 68-77.

- Retrieved from <http://wendyberrymendes.com/cms/uploads/Blascovich1999-%20Social%20facilit%282%29.pdf>
- Blass, T. (1991). Understanding behavior in the Milgram obedience experiment: The role of personality, situations, and their interactions. *Journal of Personality and Social Psychology*, 60(3), 398-413. Retrieved from <http://www.stanleymilgram.com/pdf/understanding20%behaviour.pdf>
- Blass, T. (1996). Attribution of responsibility and trust in the Milgram obedience experiment. *Journal of Applied Social Psychology*, 26, 1529-1535. doi:10.1111%2Fj.1559-1816.1996.tb00084.x
- Blass, T. (2012). A Cross-Cultural Comparison of Studies of Obedience Using the Milgram Paradigm: A Review. *Social and Personality Psychology Compass*, 6(2), 196-205. doi:10.1111/j.1751-9004.2011.00417.x
- Bock, D. C. (1972). Obedience: A response to authority and Christian commitment. *Dissertation Abstracts International*, 33, 3278-3279.
- Bock, D. C., & Warren, N. C. (1972). Religious belief as a factor in obedience to destructive commands. *Review of Religious Research*, 13, 185-191. doi:10.2307/3510781
- Bond, C. F., & Titus, L. J. (1983). Social facilitation: a meta-analysis of 241 studies. *Psychological Bulletin*, 94(2), 265-292. doi:10.1037/0033-2909.94.2.265
- Bornemann, B., Kok, B. E., Böckler, A., & Singer, T. (2016). Helping from the heart: Voluntary up-regulation of heart rate variability predicts altruistic behavior. *Biological Psychology*, 119, 54-63. doi:10.1016/j.biopsycho.2016.07.004
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652. doi:10.1037/0033-295X.108.3.624
- Bouchard, T. J. (1994). Genes, environment, and personality. *Science*, 264, 1700-1701. doi:10.1126/science.8209250
- Bouchard, T. J. (2009). Authoritarianism, Religiousness, and Conservatism: Is “Obedience to Authority” the Explanation for Their Clustering, Universality and Evolution? In E. Voland & W. Schiefenhövel (Eds.), *The Biological Evolution of Religious Mind and Behavior* (pp. 165-180). Berlin: Springer-Verlag.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago, IL: The University of Chicago Press.

- Bregman, N. J., & McAllister, H. A. (1982). Motivation and Skin Temperature Biofeedback: Yerkes-Dodson Revisited. *Psychophysiology*, *19*(3), 282-285. doi:10.1111/j.1469-8986.1982.tb02564.x
- Bridgett, D. J., Oddi, K. B., Laake, L. M., Murdock, K. W., & Bachmann, M. N. (2013). Integrating and differentiating aspects of self-regulation: effortful control, executive functioning, and links to negative affectivity. *Emotion*, *13*(1), 47-63. doi:10.1037/a0029536
- Brink, T., Urton, K., Held, D., Kirilina, E., Hofmann, M., Klann-Delius, G., Jacobs, A. M., & Kuchinke, L. (2011). The role of orbitofrontal cortex in processing empathy stories in 4-8 year-old children. *Frontiers in Psychology*, *2*, 80. doi:10.3389/fpsyg.2011.0008
- Brock, T. C., & Becker, L. A. (1966). "Debriefing" and susceptibility to subsequent experimental manipulations. *Journal of Experimental Social Psychology*, *2*, 314-323. doi:10.1016/0022-1031(66)90087-4
- Brosschot, J. F., Verkuil, B., & Thayer, J. F. (2016). The default response to uncertainty and the importance of perceived safety in anxiety and stress: an evolution-theoretical perspective. *Journal of Anxiety Disorders*, *41*, 22-34. doi:10.1016/j.janxdis.2016.04.012
- Brown, S. L., & Brown, R. M. (2015). Connecting prosocial behavior to improved physical health: Contributions from the neurobiology of parenting. *Neuroscience and Biobehavioral Reviews*, *55*, 1-17. doi:10.1016/j.neubiorev.2015.04.004
- Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist*, *64*, 1-11. doi:10.1037/a0010932
- Burger, J., Girgis, Z., & Manning, C. (2011). In Their Own Words: Explaining Obedience to Authority Through an Examination of Participants' Comments. *Social Psychological and Personality Science*, *2*(5), 460-466. doi:10.1177/1948550610397632
- Burley, P. M., & McGuiness, J. (1977). Effects of social intelligence on the Milgram paradigm. *Psychological Reports*, *40*, 767-770.
- Cacioppo, J. T., & Tassinary, L. G. (1990). Psychophysiology and psychophysiological inference. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 3-33). New York: Cambridge University Press.
- Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion Changes the Sense of Agency in the Human Brain. *Current Biology*, *26*(5), 585-592. doi:10.1016/j.cub.2015.12.067

- Cavalli-Sforza, L., & Feldman, M. (1981). *Cultural transmission and evolution: a quantitative approach*. Princeton, NJ: Princeton University Press.
- Charlesworth, W. R. (2003). *Profiling terrorists: A taxonomy of evolutionary, developmental and situational causes of a terrorist act*. *Defense & Security Analysis*, 19(3), 241-264. doi:10.1080/1475179032000118812
- Charney, E. (2008). Genes and ideologies. *Perspectives on Politics*, 6, 299-319. doi:10.1017/S1537592708080626
- Cheetham, M., Pedroni, A. F., Antley, A., Slater, M., & Jäncke, L. (2009). Virtual milgram: Empathic concern or personal distress? Evidence from functional MRI and dispositional measures. *Frontiers in Human Neuroscience*, 3(29). doi:10.3389/neuro.09.029.2009
- Chiao, J. Y., & Blizinsky, K. D. (2009). Culture-gene coevolution of individualism – collectivism and the serotonin transporter gene. *Proceedings of the Royal Society B*, 277, 529-537. doi:10.1098/rspb.2009.1650
- Chiao, J. Y., Mathur, V. A., Harada, T., & Lipke, T. (2009). Neural Basis of Preference for Human Social Hierarchy versus Egalitarianism. *Values, Empathy, and Fairness across Social Barriers*, 1167, 174-181. doi:10.1111/j.1749-6632.2009.04508.x
- Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2, 84-92. doi:10.1093/scan/nsm001
- Cook, T. D., Bean, J. R., Calder, B. J., Frey, R., Krovetz, M. L., & Reisman, S. R. (1970). Demand characteristics and three conceptions of the frequently deceived subject. *Journal of Personality and Social Psychology*, 14, 185-194.
- Cook, T. D., & Perrin, B. F. (1971). The effects of suspiciousness of deception and the perceived legitimacy of deception on task performance in an attitude change experiment. *Journal of Personality*, 39(2), 204-224. doi:10.1111/j.1467-6494.1971.tb00037.x
- Cooper, L. A. (1976). Demonstration of a mental analog of an external rotation. *Perception and Psychophysics*, 19(4), 296-302. doi:10.3758/BF03204234
- Cornelis, I., & Van Hiel, A. (2006). The impact of cognitive styles on authoritarianism based conservatism and racism. *Basic and Applied Social Psychology*, 28, 37-50. doi:10.1207/s15324834basp2801_4

- Costa, P. T., & McCrae, R. R. (1995). Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, *64*(1), 21-50. doi:10.1207/s15327752jpa6401_2
- Critchley, H. D., Elliott, R., Mathias, C. J., & Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, *20*(8), 3033-3040. Retrieved from <http://www.jneurosci.org/content/20/8/3033.long>
- Crockett, M. J. (2009). The neurochemistry of fairness: clarifying the link between serotonin and prosocial behavior. *Annals of New York Academy of Sciences*, *1167*, 76-86. doi:10.1111/j.1749-6632.2009.04506.x
- Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363-366. doi:10.1016/j.tics.2013.06.005
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*(40), 17433-17438. doi:10.1073/pnas.1009396107
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, *15*, 806-813. doi:10.1111/j.0956-7976.2004.00760.x
- Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain*. New York: Grosset/Putnam.
- Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, *41*(2), 81-94. doi:0166-4328(90)90144-4
- Dambrun, M., & Vatiné, E. (2010). Reopening the study of extreme social behaviors: Obedience to authority within an immersive video environment. *European Journal of Social Psychology*, *40*(5), 760-773. doi:10.1002/ejsp.646
- Dang, J. (2017). An updated meta-analysis of the ego-depletion effect. *Psychological Research*, 1-7. doi:10.1007/s00426-017-0862-x
- Darwin, C. R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.
- Darwin, C. R. (1871). *The descent of man, and selection in relation to sex*. London: John Murray.

- Davidson, R. J., Putnam, K. M., & Larson, C. L. (2000). Dysfunction in the neural circuitry of emotion regulation – a possible prelude to violence. *Science*, 289(5479), 591-594. doi:10.1126/science.289.5479.591
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126. doi:10.1037/0022-3514.44.1.113
- Davis, M. H. (1994). *Empathy: A social psychological approach*. Boulder, CO: Westview Press.
- Decety, J. (2010). The Neurodevelopment of Empathy in Humans. *Developmental Neuroscience*, 32(4), 257-267. doi:10.1159/000317771
- Decety, J. (2011). Dissecting the neural mechanisms mediating empathy. *Emotion Review*, 3(1), 92-108. doi:10.1177/1754073910374662
- Decety, J., Echos, S., & Correll, J. (2009). The blame game: the effect of responsibility and social stigma on empathy for pain. *Journal of Cognitive Neuroscience*, 22(5), 985-997. doi:10.1162/jocn.2009.21266
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews*, 3(2), 71-100. doi:10.1177/1534582304267187
- Decety, J., & Lamm, C. (2009). Empathy versus Personal Distress: Recent Evidence from Social Neuroscience. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 199-213). Cambridge, MA: MIT Press.
- Decety, J., & Svetlova, M. (2012). Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental Cognitive Neuroscience*, 2(1), 1-24. doi:10.1016/j.dcn.2011.05.003
- Delgado, M. R., Nearing, K. I., LeDoux, J. E., & Phelps, E. A. (2008). Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron*, 59, 829-838. doi:10.1016/j.neuron.2008.06.029
- Derryberry, D., Reed, M. A., Pilkenton-Taylor, C. (2003). Temperament and coping: advantages of an individual differences perspective. *Development and Psychopathology*, 15(4), 1049-1066.
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature*, 427, 311-312. doi:10.1038/427311a

- Dru, V. (2007). Authoritarianism, Social Dominance Orientation and Prejudice: Effects of Various Self-Categorization Conditions. *Journal of Experimental Social Psychology*, 43(6), 877-883. doi:10.1016/j.jesp.2006.10.008
- de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Science*, 10(10), 435-441. doi:10.1016/j.tics.2006.08.008
- de Waal, F. B. M. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press.
- de Waal, F. B. M. (1998[1982]). *Chimpanzee Politics: Power and Sex among Apes*. Baltimore, MD: Johns Hopkins University Press.
- de Waal, F. B. M. (2003). Morality and the Social Instincts: Continuity with the Other Primates. In G. B. Peterson (Ed.), *The Tanner Lectures on Human Values* (Vol. 25, pp. 1-40). Salt Lake City, UT: University of Utah Press. Retrieved from http://tannerlectures.utah.edu/_documents/a-to-z/d/deWaal_2005.pdf
- de Waal, F. B. M. (2008). Putting the altruism back into altruism: the evolution of empathy. *Annual Review of Psychology*, 59, 279-300. doi:10.1146/annurev.psych.59.103006.093625
- de Wied, M., van Boxtel, A., Zaalberg, R., Goudena, P. P., & Matthys, W. (2006). Facial EMG responses to dynamic emotional facial expressions in boys with disruptive behavior disorders. *Journal of Psychiatric Research*, 40 (2), 112-121. doi:10.1016/j.jpsychires.2005.08.003
- DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33, 533-552. doi:10.1016/S0191-8869(01)00171-4
- Dickman, S. J. (2002). Dimensions of Arousal: Wakefulness and Vigor. *Human Factors*, 44(3), 429-442. doi:10.1518/0018720024497673
- Doi, H., Nishitani, S., & Shinohara, K. (2013). NIRS as a tool for assaying emotional function in the prefrontal cortex. *Frontiers in Human Neuroscience*, 7, 770. doi:10.3389/fnhum.2013.00770
- Doty, R. G., Winter, D. G., Peterson, B. E., & Kimmelmeier, M. (1997). Authoritarianism and American students' attitudes about the Gulf War, 1990-1996. *Personality and Social Psychology Bulletin*, 23, 1133-1143. doi:10.1177/01461672972311002
- Dovidio, J. F., Pearson, A. R., & Orr, P. (2008). Social psychology and neuroscience: Strange bedfellows or happy marriage? *Group Processes and Intergroup Relations*, 11, 249-265. doi:10.1177/1368430207088041

- Drake, K. E. (2010). Interrogative suggestibility: Life adversity, neuroticism, and compliance. *Personality and Individual Differences, 48*, 493-498, doi:10.1016/j.paid.2009.11.030
- Duckitt, J. (1989). Authoritarianism and group identification: A new view of an old construct. *Political Psychology, 10*, 63-84. doi:10.2307/3791588
- Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 33, pp. 41-113). New York: Academic Press.
- Duckitt, J. (2013). Introduction to the special section on authoritarianism in societal context: The role of threat. *International Journal of Psychology, 48*(1), 1-5. doi:10.1080/00207594.2012.738298
- Duckitt, J., & Fisher, K. (2003). The impact of social threat on worldview and ideological attitudes. *Political Psychology, 24*(1), 199-222. doi:10.1111/0162-895X.00322
- Duckitt, J., & Sibley, C. G. (2010). Personality, Ideology, Prejudice, and Politics: A Dual-Process Motivational Model. *Journal of Personality, 78*(6), 1861-1894. doi:10.1111/j.1467-6494.2010.00672.x
- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6*(5), 178-190. doi:10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8
- Duriez, B., & Van Hiel, A. (2002). The march of modern fascism. A comparison of social dominance orientation and authoritarianism. *Personality and Individual Differences, 32*, 1199-1213. doi:10.1016/S0191-8869(01)00086-1
- Eberhardt, J. L. (2005). Imaging race. *American Psychologist, 60*(2), 181-190. doi:10.1037/0003-066X.60.2.181
- Edwards, D. M., Franks, P., Friedgood, D., Lobban, G., & Mackay, H. G. C. (1969). *An experiment on obedience*. Unpublished document, University of Witwatersand, Johannesburg, South Africa.
- Eippert, F., Veit, R., Weiskopf, N., Erb, M., Birbaumer, N., & Anders, S. (2007). Regulation of emotional responses elicited by threat-related stimuli. *Human Brain Mapping, 28*(5), 409-423. doi:10.1002/hbm.20291
- Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71-83). Cambridge, MA: MIT Press.
- Eisenberg, N., & Fabes, R. A. (1992). Emotion, regulation, and the development of social competence. In M. S. Clark (Ed.), *Review of Personality and Social Psychology: Emotion and Social Behavior* (Vol. 14, pp. 119-150). Newbury Park, CA: Sage.

- Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P. et al. (1994). The relations of emotionality and regulation to dispositional and situational empathy-related responding. *Journal of Personality and Social Psychology*, *66*(4), 776-797.
- Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Smith, M., & Maszk, P. (1996). The relations of children's dispositional empathy-related responding to their emotionality, regulation, and social functioning. *Developmental Psychology*, *32*, 195-209. doi:10.1037/0012-1649.32.2.195
- Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Prosocial development. In N. Eisenberg and W. Damon (Eds.), *Handbook of child psychology. Social, emotional, and personality development* (Vol. 3, 6th ed., pp. 646-718.). New York: Wiley.
- Eisenberg, N., Valiente, C., & Champion, C. (2004). Empathy-Related Responding: Moral, Social, and Socialization Correlates. In A. G. Miller (Ed.), *The Social Psychology of Good and Evil: Understanding Our Capacity for Kindness and Cruelty* (pp. 386-415). New York: Guilford Press.
- Ekehammar, B., & Akrami, N. (2003). The relation between personality and prejudice: a variable- and a person-centred approach. *European Journal of Personality*, *17*(6), 449-464. doi:10.1002/per.494
- Elms, A. C. (1972). *Social Psychology and Social Relevance*. Boston: Little, Brown.
- Elms, A. C. (1995). Obedience in Retrospect. *Journal of Social Issues*, *51*(3), 21-31. doi:10.1111/j.1540-4560.1995.tb01332.x
- Elms, A. C., & Milgram, S. (1966). Personality characteristics associated with obedience and defiance toward authoritative command. *Journal of Experimental Research in Personality*, *1*(4), 282-289.
- Englis, B. G., Yaughan, K. B., & Lanzetta, J. T. (1982). Conditioning of counter-empathetic emotional responses. *Journal of Experimental Social Psychology*, *18*, 375-391.
- Ertl, M., Hildebrandt, M., Ourina, K., Leicht, G., & Mulert, C. (2013). Emotion regulation by cognitive reappraisal - the role of frontal theta oscillations. *Neuroimage*, *81*, 412-421. doi:10.1016/j.neuroimage
- Eslinger, P. J. (1998). Neurological and Neuropsychological Bases of Empathy. *European Neurology*, *39*, 193-199. doi:10.1159/000007933
- Eysenck, H. J. (1955). Cortical inhibition, figural after-effect and theory of personality. *The Journal of Abnormal and Social Psychology*, *51*(1), 94-106. doi:10.1037/h0043564
- Fabes, R. A., & Eisenberg, N. (1997). Regulatory control and adults' stress-related responses to daily life events. *Journal of Personality and Social Psychology*, *73*(5), 1107-1117.

- Retrieved from
http://www.researchgate.net/profile/Nancy_Eisenberg/publication/13864629_Regulatory_control_and_adults%27_stress_related_responses_to_daily_life_events/links/02e7e539be941aade6000000.pdf
- Fabes, R. A., Eisenberg, N., & Eisenbud, L. (1993). Behavioral and Physiological Correlates of Children's Reactions to Others in Distress. *Developmental Psychology*, 29(4), 655-663.
- Feigl, H., (1958). The "Mental" and the "Physical". *Minnesota Studies in the Philosophy of Science*, 2, 370-497.
- Feldman, R. (2012). Oxytocin and social affiliation in humans. *Hormones and Behavior*, 61(3), 380-391. doi:10.1016/j.yhbeh.2012.01.008
- Feldman, S., Lavine, H., Lodge, M., & Verhulst, B. (2010). *Seeing Negative: Authoritarianism and Automatic Vigilance for Threatening Stimuli*. Paper presented at the Annual Meeting of the International Society of Political Psychology, San Francisco, CA.
- Feldman, S., & Stenner, K. (1997). Perceived Threat and Authoritarianism. *Political Psychology*, 18(4), 741-770. doi:10.1111/0162-895X.00077
- Finney, P. D. (1987). When consent information refers to risk and deception: Implications for social research. *Journal of Social Behavior and Personality*, 2(1), 37-48.
- Fischer, C. T. (1968). Ethical issues in the use of human subjects. *American Psychologist*, 23, 532. doi:10.1037/h0037714
- Fiske, A. P. (1991). *Structures of Social Life: The Four Elementary Forms of Human Relations*. New York: Free Press.
- Francis, Z., & Inzlicht, M. (2016). Proximate and ultimate causes of ego depletion. In E. Hirt (Ed.), *Self-Regulation and Ego Control* (pp.373-398). New York: Elsevier.
- Freedman, J. L. (1969). Role Playing: Psychology by Consensus. *Journal of Personality and Social Psychology*, 13(2), 107-114. doi:10.1037/h0028067
- Freeman, W. J. (2005). Origin, structure, and role of background EEG activity. Part 3. Neural frame classification. *Clinical Neurophysiology*, 116, 1118-1129. Retrieved from http://sulcus.berkeley.edu/wjf/ei_eegpart3frameclassific.pdf
- Freeman, W. J., Holmes, M. D., Burke, B. C., & Vanhatalo, S. (2003). Spatial spectra of scalp EEG and EMG from awake humans. *Clinical Neurophysiology*, 114, 1053-1068. doi:10.1016/S1388-2457(03)00045-2

- Funder, D. C., Guillaume, E., Kumagai, S., Kawamoto, S., & Sato, T. (2012). The person-situation debate and the assessment of situations. *Japanese Journal of Personality*, 21(1), 1-11. doi:10.2132/personality.21.1
- Funk, C. M., & Gazzaniga, M. S. (2009). The functional brain architecture of human morality. *Current Opinion in Neurobiology*, 19(6), 678-681. doi:10.1016/j.conb.2009.09.011
- Gambino, R. (1973). Watergate lingo: A language of non-responsibility. *Freedom at Issue*, 22(7-9), 15-17.
- Gamez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, D. (2011). Development of a measure of experiential avoidance: The Multidimensional Experiential Avoidance Questionnaire. *Psychological Assessment*, 23(3), 692-713. doi:10.1037/a0023242
- Gatto, J. & Dambrun, M. (2010). Autoritarisme et préjugés dans la police : l'effet d'une position d'infériorité numérique menaçante et du contexte normatif dans une situation de contact intergroupe. *Revue Internationale de Psychologie Sociale*, 23(4), 123-158.
- Geller, D. M. (1978). Involvement in role-playing simulations: A demonstration with studies on obedience. *Journal of Personality and Social Psychology*, 36(3), 219-235. doi:10.1037/0022-3514.36.3.219
- Gevirtz, R. (2013). The promise of heart rate variability biofeedback: evidence-based applications. *Biofeedback*, 41, 110-120. doi:10.5298/1081-5937-41.3.01
- Gilovich, T., Keltner, D., & Nisbett, R. E. (2006). *Social Psychology*. New York: W.W. Norton.
- Golding, S. L., & Lichtenstein, E. (1970). Confession of awareness and prior knowledge of deception as a function of interview set and approval motivation. *Journal of Personality and Social Psychology*, 14, 213-223. doi:10.1037/h0028853
- Goldin, P. R., Manber, T., Hakimi, S., Canli, T., & Gross, J. J. (2009). Neural bases of social anxiety disorder: Emotional reactivity and cognitive regulation during social and physical threat. *Archives of General Psychiatry*, 66, 170-180. doi:10.1001/archgenpsychiatry.2008.525
- Goubert, L., Craig, K. D., & Buysse, A. (2009). Perceiving others in pain: experimental and clinical evidence on the role of empathy. In J. Decety & W. Ickes (Eds.), *The Social Neuroscience of Empathy* (pp. 153-166), Cambridge, MA: MIT Press.
- Goubert, L., Craig, K. D., Vervoort, T., Morley, S., Sullivan, M. J. L., Williams, A. C. de C., Cano, A., & Crombez, G. (2005). Facing others in pain: The effects of empathy. *Pain*, 118, 285-288. doi:10.1016/j.pain.2005.10.025

- Goubert, L., Vervoort, T., Sullivan, M. J. L., Verhoeven, K., & Crombez, G. (2008). Parental emotional responses to their child's pain: the role of dispositional empathy and parental catastrophizing about their child's pain. *Journal of Pain, 9*, 272-279. doi:10.1016/j.jpain.2007.11.006
- Gratz, K. L., & Roemer, L. (2004). Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment, 36*, 41-54. doi:10.1007/s10862-008-9102-4
- Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A., Veeder, M., Kirkland, S., & Lyon, D. (1990). Evidence for terror management theory II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview, *Journal of Personality and Social Psychology, 58*(2), 308-318. doi:10.1037/0022-3514.58.2.308
- Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology, 45*(3), 581-584. doi:10.1016/j.jesp.2009.01.003
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389-400. doi:10.1016/j.neuron.2004.09.027
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105-2108. doi: 10.1126/science.1062872
- Greenwood, J. D. (1989). *Explanation and Experiment in Social Psychological Science*. New York: Springer-Verlag.
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271-299. doi:10.1037/1089-2680.2.3.271
- Gross, J. J. (2013). Emotion regulation: Taking stock and moving forward. *Emotion, 13*, 359-365. doi:10.1037/a0032135
- Grossman, P., & Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. *Biological Psychology, 74*(2), 263-285. doi:10.1016/j.biopsycho.2005.11.014
- Guimond, S., Dambrun, M., Michinov, N., & Duarte, S. (2003). Does social dominance generate prejudice? Integrating individual and contextual determinants of intergroup

- cognitions. *Journal of Personality and Social Psychology*, 84(4), 697-721. doi:10.1037/0022-3514.84.4.697
- Gyurak, A., Goodkind, M. S., Madan, A., Kramer, J. H., Miller, B. L., & Levenson, R. W. (2009). Do tests of executive functioning predict ability to downregulate emotions spontaneously and when instructed to suppress? *Cognitive, Affective & Behavioral Neuroscience*, 9(2), 144-152. doi:10.3758/CABN.9.2.144
- Haas, K. (1966). Obedience: Submission to destructive orders as related to hostility. *Psychological Reports*, 19, 32-34. doi:10.2466/pr0.1966.19.1.32
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5(4), 382-385. doi:10.1038/nn827
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98-116. doi:0.1007/s11211-007-0034-z
- Hallman, D. M., Olsson, E. M. G., von Schéele, B., Melin, L., & Lyskov, E. (2011). Effects of Heart Rate Variability Biofeedback in Subjects with Stress-Related Chronic Neck Pain: A Pilot Study. *Applied Psychophysiology and Biofeedback*, 36(2), 71-80. doi:10.1007/s10484-011-9147-0
- Harcourt, A. H. (1988). Alliances in contests and social intelligence. In R. Byrne & A. Whiten (Eds.), *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans* (pp. 132-152). Oxford: Clarendon Press.
- Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology*, 4, 217-246. doi:10.1146/annurev.clinpsy.3.022806.091452
- Hariri, A. R., Goldberg, T. E., Mattay, V. S., Kolachana, B. S., Callicott, J. H., Egan, M. F., & Weinberger, D. R. (2003). Brain-derived neurotrophic factor val66met polymorphism affects human memory-related hippocampal activity and predicts memory performance. *The Journal of Neuroscience*, 23(17), 6690-6694. Retrieved from <http://www.jneurosci.org/content/23/17/6690.long>
- Harris, J. A., Rushton, J., Hampson, E., & Jackson, D. N. (1996). Salivary testosterone and self-report aggressive and prosocial personality characteristics in men and women. *Aggressive Behavior*, 22, 321-331. doi:10.1002/(SICI)1098-2337(1996)22:5<321::AID-AB1>3.0.CO;2-M

- Harrison, N. A., Morgan, R., & Critchley, H. (2010). From facial mimicry to emotional empathy: A role for norepinephrine? *Social Neuroscience*, 5(4), 393-400. doi:10.1002/(SICI)1098-2337(1996)22:5<321::AID-AB1>3.0.CO;2-M
- Hart, A. J., et al. (2000). Differential response in the human amygdala to racial outgroup versus ingroup face stimuli. *Neuroreport*, 11(11), 2351-2355.
- Haslam, N., Loughnan, S., & Perry, G. (2014). Meta-Milgram: An Empirical Synthesis of the Obedience Experiments. *Plos One*, 9(4), e93927. doi:10.1371/journal.pone.0093927
- Hatemi, P. K., & McDermott, R. (2012). The genetics of politics: discovery, challenges, and progress. *Trends in Genetics*, 28(10), 525-533. doi:10.1016/j.tig.2012.07.004
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. New York: Cambridge University Press.
- Haug, T. T., Svebak, S., Hausken, T., Wilhelmsen, I., Berstad, A., & Ursin, H. (1994). Low vagal activity as mediating mechanism for the relationship between personality factors and gastric symptoms in functional dyspepsia. *Psychosomatic Medicine*, 56, 181-186. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.505.983&rep=rep1&type=pdf>
- Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change*. New York: Guilford Press.
- Heaven, P. C. L., Organ, L., Supavadeeprasis, S., & Leeson, P. (2006). War and prejudice: A study in social values, right-wing authoritarianism, and social dominance orientation. *Personality & Individual Differences*, 40, 599-608. doi:10.1016/j.paid.2005.08.005
- Hein, G., & Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology*, 18(2), 153-158. doi:10.1016/j.conb.2008.07.012
- Held, R. M., & Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments*, 1(1), 109-112.
- Herbert, S. (1966[1886]). The Factors of Organic Evolution. In H. Spencer (Ed.), *Essays: Scientific, Political and Speculative*. Osnabruck: Otto Zeller.
- Hermans, E. J., Putman, P., & van Honk, J. (2006). Testosterone administration reduces empathetic behavior: a facial mimicry study. *Psychoneuroendocrinology*, 31(7), 859-866. doi:10.1016/j.psyneuen.2006.04.002

- Hermans, E. J., van Wingen, G., Bos, P. A., Putman, P., & van Honk, J. (2009). Reduced spontaneous facial mimicry in women with autistic traits. *Biological Psychology*, *80*, 348-353. doi:10.1016/j.biopsycho.2008.12.002
- Hertwig, R., & Ortmann, A. (2008). Deception in Experiments: Revisiting the Arguments in Its Defense. *Ethics & Behavior*, *18*(1), 59-92. doi:10.1080/10508420701712990
- Hess, A. (1971). *An analysis of the Milgram (1963) study using the acting experiment*, Unpublished document, University of Nevada, Nevada.
- Hess, U., Philippot, P., & Blairy, S. (1999). Mimicry: facts and fiction. In P. Philippot, R. Feldman, & E. J. Coats (Eds.), *The Social Context of Nonverbal Behavior, Studies in Emotion and Social Interaction* (pp. 213-241). New York: Cambridge University Press.
- Hey, J. D. (1998). Experimental economics and deception. *Journal of Economic Psychology*, *19*, 397-401.
- Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences*, *37*(3), 297-350 doi:10.1017/S0140525X13001192
- Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229-1243. doi:10.1162/jocn.2009.21189
- Hobbes, T. (1991 [1651]). *Leviathan*. Cambridge: Cambridge University Press.
- Hoffman, M. L. (2002). How automatic and representational is empathy, and why. *Behavioral and Brain Sciences*, *25*, 38-39. doi:10.1017/S0140525X02410011
- Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology*, *43*(3), 497-504. doi:10.1016/j.jesp.2006.05.004
- Høgh-Olesen, H. (2010). Homo sapiens - Homo socius: A comparative analysis of human mind and kind. In H. Høgh-Olesen (Ed.), *Human morality and sociality: Evolutionary and comparative perspectives* (pp. 235-271). Basingstoke: Palgrave Macmillan.
- Holland, C. H. (1969). Sources of variance in the experimental investigation of behavioral obedience. *Dissertation Abstracts*, *29*, 2802A. (University Microfilms No. 69-2146).
- House, J. S., Landis, K. R., & Umberson, D. (1988). Social Relationships and Health. *Science*, *241*(4865), 540-545. doi:10.1126/science.3399889

- Huang, L., & Liu, J. H. (2005). Personality and social structural implications of the situational priming of social dominance orientation. *Personality and Individual Differences*, 38(2), 267-276. doi:10.1016/j.paid.2004.04.006
- Huxley, T. H. (1989 [1894]). *Evolution and Ethics*. Princeton: Princeton University Press.
- Inzlicht, M., Berkman, E., & Elkins-Brown, N. (2016). The Neuroscience of “Ego Depletion” or: How the Brain Can Help Us Understand Why Self-Control Seems Limited. In E. Harmon-Jones & M. Inzlicht (Eds.), *Social Neuroscience: Biological Approaches to Social Psychology* (pp. 101-123). New York: Routledge.
- Inzlicht, M., & Gutsell, J. N. (2007). Running on empty: Neural signals for self-control failure. *Psychological Science*, 18(11), 933-937. doi:10.1111/j.1467-9280.2007.02004.x
- Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why Self-Control Seems (But May Not Be) Limited. *Trends in Cognitive Sciences*, 18(3), 127-133. doi:10.1016/j.tics.2013.12.009
- Jackson, J. H. (1958). Evolution and dissolution of the nervous system. In J. Taylor (Ed.), *Selected writings of John Hughlings Jackson* (Vol. 2, pp. 3-92). New York: Basic Books.
- Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia*, 44(5), 752-761. doi:10.1016/j.neuropsychologia.2005.07.01
- Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, 24, 771-779. doi:10.1016/j.neuroimage.2004.09.006
- Jackson, P. L., Rainville, P., & Decety, J. (2006). To what extent do we share the pain of others? Insight from the neural bases of pain empathy. *Pain*, 125, 5-9. doi:10.1016/j.pain.2006.09.013
- Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language*, 23(2), 190-223. doi:10.1111/j.1468-0017.2007.00337.x
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102-138). New York: Guilford Press.
- Johnson, J. A. (1987). Dominance rank in juvenile olive baboons, *Papio anubis*: The influence of gender, size, maternal rank and orphaning. *Animal Behaviour*, 35, 1694-1708. doi:10.1016/S0003-3472(87)80062-3

- Johnson, D. D., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution*, 28(8), 474-481. doi:10.1016/j.tree.2013.05.014
- Jost, J. T., & Amodio, D. M. (2012). Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion*, 36(1), 55-64. doi:10.1007/s11031-011-9260-7
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3), 339-375. doi:10.1037/0033-2909.129.3.339
- Jugert, P., & Duckitt J. (2009). A Motivational Model of Authoritarianism: Integrating Personal and Situational Determinants. *Political Psychology*, 30(5), 693-719. doi:10.1111/j.1467-9221.2009.00722.x
- Jugert, P., & Hiemisch, A. (2005). Terrorismusbedrohung und sozial-kognitive Motive als Prädiktoren von konservativen Einstellungen. *Zeitschrift für Politische Psychologie*, 13, 137-158.
- Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, 134, 193-209. doi:10.1016/j.cognition.2014.10.005
- Kalinichenko, A. N., & Yur'eva, O. D. (2012). Assessment of human psychophysiological states based on methods for heart rate variability analysis. *Pattern Recognition and Image Analysis*, 22(4), 570-575. doi:10.1134/S1054661812040074
- Kalisch, R., Wiech, K., Critchley, H. D., Seymour, B., O'Doherty, J. P., Oakley, D. A., Allen, P., & Dolan, R. J. (2005). Anxiety reduction through detachment: subjective, physiological, and neural effects. *Journal of Cognitive Neuroscience*, 17, 874-883. doi:10.1162/0898929054021184
- Kanai, R., Feilden, T., Firth, C., & Rees, G. (2011). Political orientations are correlated with brain structure in young adults. *Current Biology*, 21, 677-680. doi:10.1016/j.cub.2011.03.017
- Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, 49(11), 3121-3129. doi:10.1016/j.neuropsychologia.2011.07.022
- Kaufmann, H. (1967). The price of obedience and the price of knowledge. *American Psychologist*, 22, 321-322. doi:10.1037/h0037664

- Kawanaka, K. (1982). Further studies on predation by chimpanzees of the Mahale Mountains. *Primates*, 23, 364-384. doi:10.1007/BF02381320
- Kawanaka, K. (1989). Age differences in social interactions of young males in a chimpanzee unit-group at the Mahale Mountains National Park, Tanzania. *Primates*, 30, 285-305. doi:10.1007/BF02381256
- Kelman, H. C. (1973). Violence without Moral Restraint: Reflections on the Dehumanization of Victims and Victimizers. *Journal of Social Issues*, 29(4), 25-61. doi:10.1111/j.1540-4560.1973.tb00102.x
- Kelman, H. C. (2001). Reflections on social and psychological processes of legitimization and delegitimization. In J. T. Jost & B. Major (Eds.), *The psychology of legitimacy: Emerging perspectives on ideology, justice, and intergroup relations* (pp. 54–73). Cambridge: Cambridge University Press
- Kelman, H. C. (2005). The policy context of torture: A social-psychological analysis. *International Review of the Red Cross*, 87(857), 123-134. Retrieved from <http://scholar.harvard.edu/files/hckelman/files/PolicyContextTorture.pdf?m=1360038944>
- Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of Obedience: Toward a Social Psychology of Authority and Responsibility*. New Haven: Yale University Press.
- Kilham, W., & Mann, L. (1974). Level of destructive obedience as a function of transmitter and executant roles in the Milgram obedience paradigm. *Journal of Personality and Social Psychology*, 29, 696-702. doi:10.1037/h0036636
- Kimmel, A. J. (1996). *Ethical issues in behavioral research: A survey*. Cambridge: Blackwell Publishers. doi:10.1037/0003-066X
- Kimmel, A. J. (1998). In defense of deception. *American Psychologist*, 53, 803-805. doi:10.1037/0003-066X.53.7.803
- King, M. B., & Hunt, R. A. (1972). Measuring the Religious Variable: Replication. *Journal for the Scientific Study of Religion*, 11(3), 240-251. doi:10.2307/1384548
- Knyazev, G. G., Bocharov, A. V., Slobodskaya, H. R., & Ryabichenko, T. I. (2008). Personality-linked biases in perception of emotional facial expressions. *Personality and Individual Differences*, 44(5), 1093-1104. doi:10.1016/j.paid.2007.11.001
- Koenig, L. B., & Bouchard, T. J. (2006). Genetic and environmental influences on the traditional moral values triad – authoritarianism, conservatism, and religiousness – as assessed by quantitative behavior genetic methods. In P. McNamara (Ed.), *Where god*

- and science meet: How brain and evolutionary studies alter our understanding of religion* (pp. 31-60). Westport, CN: Praeger.
- Koenigs, M., Young, L., Adolphs, R., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908-911. doi:10.1038/nature05631
- Kok, B. E., & Fredrickson, B. L. (2010). Upward spirals of the heart: Autonomic flexibility, as indexed by vagal tone, reciprocally and prospectively predicts positive emotions and social connectedness. *Biological Psychology*, *85*(3), 432-436. doi:10.1016/j.biopsycho.2010.09.005
- Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, *23*(1), 4-41. doi:10.1080/02699930802619031
- Koole, S. L., Van Dillen, L. F., & Sheppes, G. (2010). The self-regulation of emotion. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (2nd ed.). New York: Guilford Press.
- Kossowska, M., & Van Hiel, A. (2003). The Relationship Between Need for Closure and Conservative Beliefs in Western and Eastern Europe. *Political Psychology*, *24*(3), 501-518. doi:10.1111/0162-895X.00338
- Krakowski, M. (2003). Violence and serotonin: influence of impulse control, affect regulation, and social functioning. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *15*(3), 294-305. doi:10.1176/appi.neuropsych.15.3.294
- Lamm, C., Batson, C. D., & Decety, J. (2007). The Neural Substrate of Human Empathy: Effects of Perspective-taking and Cognitive Appraisal. *Journal of Cognitive Neuroscience*, *19*(1), 42-58. doi:10.1162/jocn.2007.19.1.42
- Lamm, C., & Decety, J. (2008). Is the extrastriate body area (EBA) sensitive to the perception of pain in others? *Cerebral Cortex*, *18*(10), 2369-2373. doi:10.1093/cercor/bhn006.
- Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, *54*(3), 2492-2502. doi:10.1016/j.neuroimage.2010.10.014
- Lamm, C., Meltzoff, A. N., & Decety, J. (2009). How Do We Empathize with Someone Who Is Not Like Us? A Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience*, *22*(2), 362-376. doi:10.1162/jocn.2009.21186
- Lamm, C., Nusbaum, H. C., Meltzoff, A. N., & Decety, J. (2007). What are you feeling? Using functional magnetic resonance imaging to assess the modulation of sensory and affective responses during empathy for pain. *Plos One*, *2*(12), e1292. doi:10.1371/journal.pone.0001292

- Lane, R. D., Fink, G. R., Chau, P. M., & Dolan, R. J. (1997). Neural activation during selective attention to subjective emotional responses. *Neuroreport*, 8(18), 3969-3972.
- Lankford, A. (2009). *Human Killing Machines: Systematic Indoctrination in Iran, Nazi Germany, Al Qaeda, and Abu Ghraib*. Lanham, MD: Lexington Books.
- Lanzetta, J. T., & Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *Journal of Personality and Social Psychology*, 56(4), 543-554. doi:10.1037/0022-3514.56.4.543
- Lavine, H., Lodge, M., & Freitas, K. (2005). Threat, Authoritarianism, and Selective Exposure to Information. *Political Psychology*, 26(2), 219-244. doi:10.1111/j.1467-9221.2005.00416.x
- Laufs, H., Krakow, K., Sterzer, P., Eger, E., Beyerle, A., Salek-Haddadi, A., & Kleinschmidt, A. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *PNAS*, 100(16), 11053-11058. doi:10.1073/pnas.1831638100
- Lavine, H., Sweeney, D., & Wagne S. H. (1999). Depicting Women as Sex Objects in Television Advertising: Effects on Body Dissatisfaction. *Personality and Social Psychology Bulletin*, 25(8), 1049-1058. doi:10.1177/0146167299251101
- LeDoux, J. E. (1996). *The Emotional Brain*, New York: Simon & Schuster.
- Lehrer, P. M., & Gevirtz, R. (2014). Heart rate variability biofeedback: how and why does it work? *Frontiers in Psychology*, 5, 756. doi:10.3389/fpsyg.2014.00756
- Lehrer, P. M., Vaschillo, E., & Vaschillo, B. (2000). Resonant frequency biofeedback training to increase cardiac variability: rationale and manual for training. *Applied Psychophysiological Biofeedback*, 25(3), 177-191.
- Lehrer, P. M., Vaschillo, E., Vaschillo, B., Lu, S.-E., Scardella, A., Siddique, M., & Habib, R. H. (2004). Biofeedback treatment for asthma. *Chest*, 126 (2), 352-61. doi:10.1378/chest.126.2.352
- Leigh, S. R., & Shea, B. T. (1995). Ontogeny and the evolution of adult body size dimorphism in apes. *American Journal of Primatology*, 36, 37-60. doi:10.1002/ajp.1350360104
- Leonard, J. W. (1979). A strategy approach to the study of primate dominance behaviour. *Behavioural Processes*, 4(2), 99-196. doi:10.1016/0376-6357(79)90030-5
- Lesting, J., Narayanan, R. T., Kluge, C., Sangha, S., Seidenbecher, T., & Pape, H. C. (2011). Patterns of coupled theta activity in amygdala-hippocampal-prefrontal circuits during fear extinction. *Plos One*, 6(6), e21714. doi:10.1371/journal.pone.0021714

- Levy, L. (1967). Awareness, learning and the beneficent subject as expert witness. *Journal of Personality and Social Psychology*, 6, 363-370. doi:10.1037/h0024716
- Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, 8, 720-722.
- Locke, J. (1988[1689]). *Two Treatises of Government*. Cambridge: Cambridge University Press.
- Loomis, J., Blascovich, J., & Beall, A. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, 31, 557-564. doi:10.3758/BF03200735
- Lumsden, C. J., & Wilson, E. O. (1981) *Genes, Mind, and Culture: The Coevolutionary Process*. Cambridge, Massachusetts: Harvard University Press.
- Maguire, E. A., Woollett, K., & Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus*, 16(12), 1091-1101. doi:10.1002/hipo.20233
- Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: A clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience*, 10(3), 141-152. doi:10.1038/nrn2596
- Mantell, D. M. (1971). The potential for violence in Germany. *Journal of Social Issues*, 27(4), 101-112. doi:10.1111/j.1540-4560.1971.tb00680.x
- Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *Neuroimage*, 51(4), 1468-1475. doi:10.1016/j.neuroimage.2010.03.025
- Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces*, 64, 377-402. doi:10.1093/sf/64.2.377
- McCann, S. J. H. (1997). Threatening Times, 'Strong' Presidential Popular Vote Winners, and the Victory Margin, 1824-1964. *Journal of Personality and Social Psychology*, 73(1), 160-170. doi:10.1037/0022-3514.73.1.160
- McCourt, K., Bouchard, T. J., Lykken, D. T., Tellegen, A., & Keyes, M. (1999). Authoritarianism revisited: genetic and environmental influences examined in twins reared apart and together. *Personality and Individual Differences*, 27, 985-1014. doi:10.1016/S0191-8869(99)00048-3

- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81-90. doi:10.1037/0022-3514.52.1.81
- McCrae, R. R., & John O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175-215. doi:10.1111/j.1467-6494.1992.tb00970.x
- McFarland, S. G. (2005). On the eve of war: Authoritarianism, social dominance, and American students' attitudes toward attacking Iraq. *Personality and Social Psychology Bulletin*, 31(3), 360-367. doi:10.1177/0146167204271596
- Mead, N., Baumeister, R. F., Gino, F., Schweitzer, M., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, 45 (3), 594-597. doi:10.1016/j.jesp.2009.02.004
- Meeus, W. H. J., & Raaijmaker, Q. A. W. (1995). Obedience in Modern Society: The Utrecht Studies. *The Society for the Psychological Study of Social Issues*, 51(3), 155-175. doi:10.1111/j.1540-4560.1995.tb01339.x
- Mehta, P. H., & Beer, J. (2010). Neural mechanisms of the testosterone-aggression relation: the role of orbitofrontal cortex. *Journal of Cognitive Neuroscience*, 22(10), 2357-2368. doi:10.1162/jocn.2009.21389
- Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society, London: Biological Sciences*, 358, 491-500. doi:10.1098/rstb.2002.1261
- Mendez, M. F. (2006). What frontotemporal dementia reveals about the neurobiological basis of morality. *Medical Hypotheses*, 67(2), 411-418. doi:10.1016/j.mehy.2006.01.048
- Mermillod, M., & Lepage, J. (2016). Embodied Emotion: the Functional Mystery of Embodying Emotions. In Y. Coello & M. H. Fischer (Eds.), *Foundations of Embodied Cognition, Perceptual and Emotional Embodiment (Vol. 1)*. Hove: Psychology Press.
- Meyer-Lindenberg, A. (2010). From maps to mechanisms through neuroimaging of schizophrenia. *Nature*, 468(7321), 194-202. doi:10.1038/nature09569
- Miczek, K. A., de Almeida, R. M. M., Kravitz, E. A., Rissman, E. F., de Boer, S. F., & Raine, A. (2007). Neurobiology of Escalated Aggression and Violence. *The Journal of Neuroscience*, 27(44), 11803-11806. doi:10.1523/JNEUROSCI.3500-07.2007
- Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371-378. doi:10.1037/h0040525

- Milgram, S. (1965). Some Conditions of Obedience and Disobedience to Authority. *Human Relations*, 18(1), 57-76. doi:10.1177/001872676501800105
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. London: Tavistock Publications.
- Miller, F. D. (1975). *An experimental study of obedience to authority of varying legitimacy* (Doctoral dissertation). Harvard University, Cambridge, MA.
- Miller, A. G., Gillen, B., Schenker, C., & Radlove, S. (1974). The prediction and perception of obedience to authority. *Journal of Personality*, 42(1), 23-42. doi:10.1111/j.1467-6494.1974.tb00555.x
- Miller, J. G., Kahle, S., & Hastings P. D. (2015). Roots and Benefits of Costly Giving: Children Who Are More Altruistic Have Greater Autonomic Flexibility and Less Family Wealth. *Psychological Science*, 26(7), 1038-4105. doi:10.1177/0956797615578476
- Milyavskaya, M., & Inzlicht, M. (2016). Attentional and motivational mechanisms of self-control. In D. de Ridder, M. Adriaanse, & K. Fujita (Eds), *Handbook of Self-Control in Health and Well-Being*. New York: Routledge. Retrieved from <https://static1.squarespace.com/static/550b09eae4b0147d03eda40d/t/56f9cb3907eaa0601bad7717/1459211067303/attentional-and-motivational-mechanisms-of-self-control.pdf>
- Miranda, F. S. B., Caballero, R. B., Gomez, M. N. G., & Zamorano, M. A. M. (1981). Obediencia a la autoridad. *Psiquis*, 2, 212-221.
- Mixon, D. (1972). Instead of deception. *Journal for the Theory of Social Behavior*, 2, 145-177. doi:10.1111/j.1468-5914.1972.tb00309.x
- Mixon, D. (1977). Temporary false beliefs. *Personality and Social Psychology Bulletin*, 3(3), 479-488.
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Science*, 11(8), 319-321. doi:10.1016/j.tics.2007.06.001
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6(10), 799-809. doi:10.1038/nrn1768
- Morinaga, K., Akiyoshi, J., Matsushita, H., Ichioka, S., Tanaka, Y., Tsuru, J., et al. (2007). Anticipatory anxiety-induced changes in human lateral prefrontal cortex activity. *Biological Psychology*, 74, 34-38. doi:10.1016/j.biopsycho.2006.06.005

- Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, 383(6603), 812-815. doi:10.1038/383812a0
- Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences of the United States of America*, 96(4), 1680-1685.
- Morrison, I., Lloyd, D., di Pellegrino, G., & Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? *Cognitive Affective and Behavioral Neuroscience*, 4(2), 270-278. doi:10.3758/CABN.4.2.270
- Morrison, K. R., & Ybarra, O. (2008). The effects of realistic threat and group identification on social dominance orientation. *Journal of Experimental Social Psychology*, 44(1), 156-163. doi:10.1016/j.jesp.2006.12.006
- Movius, H. L., & Allen, J. J. B. (2005). Cardiac vagal tone, defensiveness, and motivational style. *Biological Psychology*, 68(2), 147-162. doi:10.1016/j.biopsycho.2004.03.019
- Moss, F.A., & Hunt, T. (1927). Are you socially intelligent? *Scientific American*, 137(2), 108-110. doi:10.1038/scientificamerican0827-108
- Muhtadie, L., Koslov, K., Akinola, M., & Mendes, W. B. (2015). Vagal flexibility: A physiological predictor of social sensitivity. *Journal of Personality and Social Psychology*, 109(1), 106-120. doi:10.1037/pspp0000016
- Murphy, B. C., Shepard, S. A., Eisenberg, N., Fabes, R. A., & Guthrie, I. K. (1999). Contemporaneous and Longitudinal Relations of Dispositional Sympathy to Emotionality, Regulation, and Social Functioning. *Journal of Early Adolescence*, 19(1), 66-97. doi:10.1177/0272431699019001004
- Myers, D. (2013). *Social Psychology (11th ed.)*. New York: McGraw-Hill.
- Nadler, R. D. (1988). Sexual Aggression in the Great Apes. *Annals of the New York Academy of Sciences*, 528, 154-162. doi:10.1111/j.1749-6632.1988.tb50858.x
- Nagai, Y., Critchley, H. D., Featherstone, E., Trimble, M. D., & Dolan, R. J. (2004). Activity in ventromedial prefrontal cortex covaries with sympathetic skin conductance level: a physiological account of a "default mode" of brain function. *Neuroimage*, 22(1), 243-251. doi:10.1016/j.neuroimage.2004.01.019
- Namiki, C., Yamada, M., Yoshida, H., Hanakawa, T., Fukuyama, H., & Murai, T. (2008). Small orbitofrontal traumatic lesions detected by high resolution MRI in a patient with major behavioural changes. *Neurocase*, 14(6), 474-479. doi:10.1080/13554790802459494

- Narayanan, V., Heiming, R. S., Jansen, F., Lesting, J., Sachser, N., Pape, H-C, et al. (2011). Social Defeat: Impact on Fear Extinction and Amygdala-Prefrontal Cortical Theta Synchrony in 5-HTT Deficient Mice. *Plos One*, 6(7), e22600. doi:10.1371/journal.pone.0022600
- Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2011). Human threat management systems: Self-protection and disease avoidance. *Neuroscience and Biobehavioral Reviews*, 35(4), 1042-1051. doi:10.1016/j.neubiorev.2010.08.011
- Neumann, D. L., & Westbury, H. R. (2011). The Psychological Measurement of Empathy. In D. J. Scapaletti (Ed), *Psychology of Empathy* (pp. 1-24). New York: Nova Science Publishers Inc.
- Niedenthal, P. M. (2007). Embodying emotion. *Science*, 316, 1002-1005. doi:10.1126/science.1136930
- Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and Brain Sciences*, 33(6), 417-433. doi:10.1017/S0140525X10000865
- Nissani, M. (1990). A cognitive reinterpretation of Stanley Milgram's observations on obedience to authority. *American Psychologist*, 45(12), 1384-1385. doi:10.1037/0003-066X.45.12.1384
- Obradović, J., Bush, N. R., Stamperdahl, J., Adler, N. E., & Boyce, W. T. (2010). Biological sensitivity to context: The interactive effects of stress reactivity and family adversity on socioemotional behavior and school readiness. *Child Development*, 81(1), 270-289. doi:10.1111/j.1467-8624.2009.01394.x
- Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. (2002). Rethinking feelings: an fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, 14(8), 1215-1229. doi:10.1162/089892902760807212
- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242-249. doi:10.1016/j.tics.2005.03.010
- Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage*, 23, 483-499. doi:10.1016/j.neuroimage.2004.06.030
- O'Leary, C., Willis, F., & Tomich, E. (1970). Conformity under deceptive and non deceptive techniques. *Sociological Quarterly*, 11(1), 87-93. doi:10.1111/j.1533-8525.1970.tb02077.x

- Opotow, S. (1990). Moral exclusion and injustice: An introduction. *Journal of Social Issues*, 46(1), 1-20. doi:10.1111/j.1540-4560.1990.tb00268.x
- Orne, M. T. (2009). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143-179). New York: Academic Press.
- Otzenberger, H., Gronfier, C., Simon, C., Charloux, A., Ehrhart, J., Piquard, F., & Brandenberger, G. (1998). Dynamic heart rate variability: a tool for exploring sympathovagal balance continuously during sleep in men. *The American Journal of Physiology*, 275(3 pt 2), H946-H950. Retrieved from <http://ajpheart.physiology.org/content/275/3/H946.long>
- Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller, J. L., Scalora, M., Hatemi, P. K., & Hibbing, J. R. (2008). Political Attitudes Vary with Physiological Traits. *Science*, 321(5896), 1667-1670. doi:10.1126/science.11576
- Packer, D. J. (2008). Identifying systematic disobedience in Milgram's obedience experiments: A meta-analytic review. *Perspectives on Psychological Science*, 3(4), 301-304. doi:10.1111/j.1745-6924.2008.00080.x
- Pagani, M., Lombardi, F., Guzzetti, S., Rimoldi, O., Furlan, R., Pizzinelli, P., Sandrone, G., Malfatto, G., Dell'Orto, S., Piccaluga, E., et al. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circulation Research*, 59(2), 178-193. doi:10.1161/01.RES.59.2.178
- Palm, K. M., Polusny, M. A., & Follette, V. M. (2004). Vicarious traumatization: potential hazards and interventions for disaster and trauma workers. *Prehospital and Disaster Medicine*, 19(1), 73-78. doi:10.1017/S1049023X00001503
- Park, G., & Thayer, J. F. (2014). From the heart to the mind: cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Frontiers in Psychology*, 5, 278. doi:10.3389/fpsyg.2014.00278
- Park, G., Vasey, M., Van Bavel, J. J., & Thayer, J. F. (2013). Cardiac vagal tone is correlated with selective attention to neutral distractors under load. *Psychophysiology*, 50(4), 398-406. doi:10.1111/psyp.12029
- Passini, S., & Morselli, D. (2009). Authority relationships between obedience and disobedience. *New Ideas in Psychology*, 27, 96-106. doi:10.1016/j.newideapsych.2008.06.001

- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*(1), 163-177. doi:10.1111/j.1551-6709.2011.01210.x.
- Payen, P. (2012). *Les revers de la guerre en Grèce ancienne. Histoire et historiographie*. Paris: Belin.
- Pertaub, D. P., Slater, M., & Barker, C. (2002). An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. *Presence*, *11*(1), 68-78. doi:10.1162/105474602317343668
- Petrovic, P., Carlsson, K., Petersson, K. M., Hansson, P., & Ingvar, M. (2004) Context-dependent deactivation of the amygdala during pain. *Journal of Cognitive Neuroscience*, *16*(7), 1289-1301. doi:10.1162/0898929041920469
- Petrovic, P., Ingvar, M., Stone-Elander, S., Petersson, K. M., & Hansson, P. (1999). A PET activation study of dynamic mechanical allodynia in patients with mononeuropathy. *Pain*, *83*(3), 459-470. doi:10.1016/S0304-3959(99)00150-5
- Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, *110*(11), 1842-1857. doi:10.1016/S1388-2457(99)00141-8
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*(5), 729-738. doi:10.1162/089892900562552
- Place, U. T. (2011[1956]). Is consciousness a brain process? *British Journal of Psychology*, *47*(1), 44-50. doi:10.1111/j.2044-8295.1956.tb00560.x
- Porges, S. W. (1992). Vagal Tone: A physiological marker of stress vulnerability. *Pediatrics*, *90*(3 pt 2), 498-504. Retrieved from http://www.researchgate.net/publication/21641337_Vagal_tone_a_physiologic_marker_of_stress_vulnerability
- Porges, S. W. (1994). Le tonus vagal: indicateur du stress et de la vulnérabilité au stress chez l'enfant. *Medecine & Enfance*, *14*, 494-503.
- Porges, S. W. (1995). Orienting in a defensive world: Mammalian modifications of our evolutionary heritage: A Polyvagal Theory. *Psychophysiology*, *32*(4), 301-318. doi:10.1111/j.1469-8986.1995.tb01213.x
- Porges, S. W. (1996). Physiological regulation in high-risk infants: A model for assessment and potential intervention. *Development and Psychopathology*, *8*(1), 43-58. doi:10.1017/S0954579400006969

- Porges, S. W. (1998). Love: An emergent property of the mammalian autonomic nervous system. *Psychoneuroendocrinology*, 23(8), 837-861. doi:10.1016/S0306-4530(98)00057-2
- Porges, S. W. (2001). The Polyvagal Theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology*, 42(2), 123-146. doi:10.1016/S0167-8760(01)00162-3
- Porges, S. W. (2003). The Polyvagal Theory: Phylogenetic contributions to social behavior. *Physiology and Behavior*, 79(3), 503-513. doi:10.1016/S0031-9384(03)00156-2
- Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, 74(2), 116-143. doi:10.1016/j.biopsycho.2006.06.009
- Posner, M. I., & Rothbart, M. K. (2007). Research on Attention Networks as a Model for the Integration of Psychological Science. *Annual Review of Psychology*, 58(1), 1-23. doi:10.1146/annurev.psych.58.110405.085516
- Pratto, F., Sidanius, J., & Levin, S. (2006). Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. *European Review of Social Psychology*, 17(1), 271-320. doi:10.1080/10463280601055772
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67(4), 741-763. doi:10.1037/0022-3514.67.4.741
- Preston S. D., & de Waal F. B. M. (2002). Empathy: its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25(1), 1-72. doi:10.1017/S0140525X02000018
- Price, D. D. (2000). Psychological and neural mechanisms of the affective dimension of pain. *Science*, 288(5472), 1769-1772. doi:10.1126/science.288.5472.1769
- Ramus, F. (2008). Influences génétiques sur le développement des fonctions cognitives. *Paper presented at the AFEP annual meeting « Précocité Intellectuelle et Apprentissages »*, Paris, France. Retrieved from <http://www.lscp.net/persons/ramus/docs/AFEP.Ramus.2008.pdf>
- Reicher, S. D., & Haslam, S. A. (2011). After shock? Towards a social identity explanation of the Milgram 'obedience' studies. *The British Journal of social psychology*, 50(Pt 1), 163-9. doi:10.1111/j.2044-8309.2010.02015.x
- Reicher, S. D., Haslam, S. A., & Smith, J. R. (2012). Working toward the experimenter: reconceptualising obedience within the Milgram paradigm as identification-based followership. *Perspectives on Psychological Science*, 7(4), 315-324. doi:10.1177/1745691612448482

- Rickert, E. J. (1998). Authoritarianism and economic threat: Implications for political behavior. *Political Psychology, 19*(4), 707-720. doi:10.1111/0162-895X.00128
- Riecken, H. W. (1962). A program for research on experiments in social psychology. In N. F. Washburne (Ed.), *Decisions, values and groups* (Vol. 2, pp. 25-41). New York: Pergamon Press.
- Rohan, M. J., & Zanna, M. P. (1996). Value transmission in families. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.), *The Ontario symposium: The psychology of values* (Vol. 8, pp. 253-276). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Rokeach, M. (1954). The nature and meaning of dogmatism. *Psychological Review, 61*(3), 194-204. doi:10.1037/h0060752
- Rokeach, M. (1960). *The open and closed mind*. New York: Basic Books.
- Rosier, M., & Willig, C. (2002). The strange death of the authoritarian personality: 50 years of psychological and political debate. *History of the Human Sciences, 15*(4), 71-96. doi:10.1177/952695102015004682
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.
- Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, D. William, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, 6th ed., pp. 99-166). Hoboken, NJ: Wiley.
- Rothbart, M. K., Derryberry, D., & Posner, M. I. (1994). A psychobiological approach to the development of temperament. In J. E. Bates & T. D. Wachs (Eds.), *Temperament: Individual differences at the interface of biology and behavior* (pp. 83-116). Washington, DC: American Psychological Association.
- Rothbart, M. K., Ellis, L. K., & Posner, M. I. (2011). Temperament and self-regulation. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory and applications* (2nd ed., pp. 441-460). New York: Guilford Press.
- Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28*(2), 573-594. doi:10.1207/s15326942dn2802_02
- Rushton, J. P., Littlefield, C. H., & Lumsden, C. J. (1986). Gene-culture coevolution of complex social behavior: human altruism and mate choice. *Proceedings of the National Academy of Sciences, 83*(19), 7340-7343. doi:10.1073/pnas.83.19.7340

- Sales, S. M. (1972). Economic Threat as a Determinant of Conversion Rates in Authoritarian and Nonauthoritarian Churches. *Journal of Personality and Social Psychology*, 23(3), 420-428. doi:10.1037/h0033157
- Sales, S. M. (1973). Threat as a factor in authoritarianism: an analysis of archival data. *Journal of Personality and Social Psychology*, 28(1), 44-57. doi:10.1037/h0035588
- Sales, S. M., & Friend, K. E. (1973). Success and Failure as Determinants of Level of Authoritarianism. *Behavioral Science*, 18(3), 163-172. doi:10.1002/bs.3830180304
- Sandi, C., & Haller, J. (2015). Stress and the social brain: Behavioural effects and neurobiological mechanisms. *Nature Reviews Neuroscience*, 16(5), 290-304. doi:10.1038/nrn3918
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332-339. doi:10.1038/nrn1651
- Sapolsky, R. M. (1993). The physiology of dominance in stable versus unstable social hierarchies. In W. A. Mason & S. P. Mendoza (Eds), *Primates social conflict* (pp. 171-204). Albany: State University of New York Press.
- Sapolsky, R. M. (1995). Social subordination as a marker of hypercortisolism: Some unexpected subtleties. In G. P. Chrousos, R. McCarty, K. Pacak, G. Cizza, E. Sternberg, P. W. Gold, & R. Kvetnansky (Eds), *Stress: Basic mechanisms and clinical implications. Annals of the New York Academy of Sciences* (Vol. 771). New York Academy of Sciences.
- Saucier, G. (1994). Separating description and evaluation in the structure of personality attributes. *Journal of Personality and Social Psychology*, 66(1), 141-154. doi:10.1037/0022-3514.66.1.141
- Saucier, G. (2000). Isms and the structure of social attitudes. *Journal of Personality and Social Psychology*, 78(2), 366-385. doi:10.1037/0022-3514.78.2.366
- Schilbach, L., Wohlschlaeger, A., Krämer, N. C., Newen, A., Zilles, K., Shah, J. N., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, 44(5), 718-730. doi:10.1016/j.neuropsychologia.2005.07.017
- Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology General*, 136(2), 241-255. doi:10.1037/0096-3445.136.2.241

- Schmitt, M. T., Branscombe, N. R., & Kappen, D. M. (2003). Attitudes toward group-based inequality: social dominance or social identity? *The British Journal of Social Psychology*, *42*(2), 161-186. doi:10.1348/014466603322127166
- Schulze, L., Domes, G., Krüger, A., Berger, C., Fleischer, M., Prehn, K., Schmahl, C., Grossmann, A., Hauenstein, K., & Herpertz, S. C. (2011). Neuronal Correlates of Cognitive Reappraisal in Borderline Patients with Affective Instability. *Biological Psychiatry*, *69*(6), 564-573. doi:10.1016/j.biopsych.2010.10.025
- Schurz, G. (1985). Experimentelle Überprüfung des Zusammenhangs zwischen Persönlichkeitsmerkmalen und der Bereitschaft zum destruktiven Gehorsam gegenüber Autoritäten. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *32*, 160-177.
- Schwartz, S. H., Caprara, G. V., & Vecchione, M. (2010). Basic Personal Values, Core Political Values, and Voting: A Longitudinal Analysis. *Political Psychology*, *31*(3), 421-452. doi:10.1111/j.1467-9221.2010.00764.x
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513-523. doi:10.1037/0022-3514.45.3.513
- Schwartz, N., & Clore, G. L. (1996). Feelings and Phenomenal Experiences. In E. Tory Higgins & A. W. Kruglanski (Eds), *Social psychology: Handbook of basic principles* (pp. 433-465). New York: Guilford.
- Schwartz, G. E., Fair, P. L., Mandel, M. R., Salt, P., Mieske, M., & Klerman, G. L. (1978). Facial electromyography in the assessment of improvement in depression. *Psychosomatic Medicine*, *40*(4), 355-360.
- Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. (1976). Facial muscle patterning to affective imagery in depressed and nondepressed subjects. *Science*, *192*(4338), 489-491.
- Segerstrom, S. C., & Nes, S. L. (2007). Heart rate variability indexes self-regulatory strength, effort, and fatigue. *Psychological Science*, *18*(3), 275-281. doi:10.1111/j.1467-9280.2007.01888.x
- Shallice, T. (1988) *From Neuropsychology to Mental Structure*. New York: Cambridge University Press.
- Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, *46*(5), 668-677. doi:10.1016/j.cortex.2009.04.008

- Shamay-Tsoory, S. G., Tomer, R., Goldsher, D., Berger, B. D., & Aharon-Peretz, J. (2004). Impairment in cognitive and affective empathy in patients with brain lesions: anatomical and cognitive correlates. *Journal of Clinical and Experimental Neuropsychology*, 26(8), 1113-1127. doi:10.1080/13803390490515531
- Shanab, M. E., & Yahya, K. A. (1978). A cross-cultural study of obedience. *Bulletin of the Psychonomic Society*, 11(4), 267-269. doi:10.3758/BF03336827
- Sheridan, T. B. (1992) Musings on Telepresence and Virtual Presence, *Telepresence. Presence: Teleoperators and Virtual Environments*, 1(1), 120-126. doi:10.1162/pres.1992.1.1.120
- Shook, N., Pena, P., Fazio, R. H., Sollers, J. J., & Thayer, J. F. (2007). Friend or foe: Heart rate variability and the negativity bias in learning about novel objects. *Psychophysiology*, 44, S39-S39.
- Shultz, S., & Dunbar, R. I. M (2007). The evolution of the social brain: anthropoid primates contrast with other vertebrates. *Proceedings of the Royal Society of London B*, 274(1624), 2429-2436. doi:10.1098/rspb.2007.0693
- Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review*, 12(3), 248-279. doi:10.1177/1088868308319226
- Sidanius, J., & Pratto, F. (1999). *Social dominance: an intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.
- Silverman, I., Shulman, A. D., & Wiesenhal, D. (1970). Effects of deception and debriefing of psychological subjects on performance in later experiments. *Journal of Personality and Social Psychology*, 14(3), 203-212. doi:10.1037/h0028852
- Simon, D., Craig, K. D., Miltner, W. H., & Rainville, P. (2006). Brain responses to dynamic facial expressions of pain. *Pain*, 126(1-3), 309-318. doi:10.1016/j.pain.2007.05.008
- Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156, 81-96. doi:10.1111/j.1749-6632.2009.04418.x
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157-1162. doi:10.1126/science.1093535
- Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, 439(7075), 466-469. doi:10.1038/nature04271

- Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *Plos One*, *1*(1), e39. doi:10.1371/journal.pone.0000039
- Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, *68*(2), 141-56. doi:10.2307/2182164
- Somit, A., & Peterson, S. A. (1997). *Darwinism, dominance and democracy: The biological bases of authoritarianism*. Westport, CT: Praeger.
- Sonnby-Borgstrom, M., Jonsson, P., & Svensson, O. (2003). Emotional empathy as related to mimicry reactions at different levels of information processing. *Journal of Nonverbal Behavior*, *27*(1), 3-23. doi:10.1023/A:1023608506243
- Stalheim, E. G., Eriksson, E., von Knorring, L., & L. Wide (1998). Testosterone as a biological marker in psychopathy and alcoholism. *Psychiatry Research*, *77*(2), 79-88. doi:10.1016/S0165-1781(97)00143-1
- Stanley, D. A., & Adolphs, R. (2013). Toward a Neural Basis for Social Behavior. *Neuron*, *80*(3), 816-826. doi:10.1016/j.neuron.2013.10.038
- Stellar, J. E., Cohen, A., Oveis, C., & Keltner, D. (2015). Affective and physiological responses to the suffering of others: Compassion and vagal activity. *Journal of Personality and Social Psychology*, *108*(4), 572-585. doi:10.1037/pspi0000010
- Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23-46). Hillsdale, NJ: Lawrence Erlbaum.
- Stone, W.F., Lederer, G., & Christie, R. (1993). *Strength and weakness: the authoritarian personality today*. New York: Springer-Verlag.
- Stössel, K., Kämpfe, N., & Riemann, R. (2006). The Jena Twin Registry and the Jena Twin Study of Social Attitudes. *Twin Research and Human Genetics*, *9*(6), 783-786. doi:10.1163/156853994X00505
- Strack, F., & Deutsch, R. (2004). Reflective and Impulsive Determinants of Social Behavior. *Personality and Social Psychology Review*, *8*(3), 220-247. doi:10.1207/s15327957pspr0803_1
- Strier, K. B. (1994). Brotherhoods among Atelins: Kinship, Affiliation, and Competition. *Behaviour*, *130*(3/4), 151-167. doi:10.1163/156853994X00505
- Sutton, R., & Douglas, K. (2013). *Social Psychology*. Basingstoke: Palgrave Macmillan.

- Swinth, K. R., & Blascovich, J. (2001). Conformity to group norms within an immersive virtual environment. *Paper presented at the annual meeting of the American Psychological Society*, Toronto, Ontario, Canada.
- Taborsky, B., & Oliveira, R. F. (2012). Social competence: an evolutionary approach. *Trends in Ecology & Evolution*, *27*(12), 679-688. doi:10.1016/j.tree.2012.09.003
- Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996). Heart rate variability. Standards of measurement, physiological interpretation and clinical use. *Circulation*, *93*(5), 1043-1065. Retrieved from http://www.escardio.org/static_file/Escardio/Guidelines/Scientific-Statements/guidelines-Heart-Rate-Variability-FT-1996.pdf
- Tassy, S., Deruelle, C., Mancini, J., Leistedt, S., & Wicker, B. (2013). High levels of psychopathic traits alters moral choice but not moral judgment. *Frontiers in Human Neuroscience*, *7*, 229. doi:10.3389/fnhum.2013.00229
- Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., Attarian, S., Felician, O., & Wicker, B. (2012). Disrupting the right prefrontal cortex alters moral judgement. *Social Cognitive & Affective Neuroscience*, *7*(3), 282-288. doi:10.1093/scan/nsr008
- Taylor, E. W., Leite, C. A. C., Sartori, M. R., Wang, T., Abe, A. S., & Crossley, D. A. (2014). The phylogeny and ontogeny of autonomic control of the heart and cardiorespiratory interactions in vertebrates. *The Journal of Experimental Biology*, *217*(5), 690-703. doi:10.1242/jeb.086199
- Thayer, J. F., Ahs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neuroscience and Biobehavioral Reviews*, *36*(2), 747-756. doi:10.1016/j.neubiorev.2011.11.009
- Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, *37*(2), 141-153. doi:10.1007/s12160-009-9101-z
- Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, *61*(3), 201-216. doi:10.1016/S0165-0327(00)00338-4
- Thayer, J. F., & Lane, R. D. (2007). The role of vagal function in the risk for cardiovascular disease and mortality. *Biological Psychology*, *74*(2), 224-242. doi:10.1016/j.biopsycho.2005.11.013

- Turner, C. W., & Simons, L. S. (1974). Effects of subject sophistication and evaluation apprehension on aggressive responses to weapons. *Journal of Personality and Social Psychology, 30*(3), 341-348. doi:10.1037/h0036894
- Uchino, B. N., Cacioppo, J. T., & Kiecolt-Glaser, J. K. (1996). The relationship between social support and physiological processes: A review with emphasis on underlying mechanisms. *Psychological Bulletin, 119*(3), 488-531. doi:10.1037/0033-2909.119.3.488
- Valiente, C., Eisenberg, N., Shepard, S. A., Fabes, R.A., Cumberland, A. J., Losoya, S. H., & Spinrad, T. L. (2004). The relations of mothers' negative expressivity to children's experience and expression of negative emotion. *Journal of Applied Developmental Psychology, 25*(2), 215-235. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2898161/>
- Van Hiel, A., & Mervielde, I. (2002). Explaining Conservative Beliefs and Political Preferences: A Comparison of Social Dominance Orientation and Authoritarianism. *Journal of Applied Social Psychology, 32*(5), 965-976. doi:10.1111/j.1559-1816.2002.tb00250.x
- Van Hiel, A., Pandelaere, M., & Duriez, B. (2004). The Impact of Need for Closure on Conservative Beliefs and Racism: Differential Mediation by Authoritarian Submission and Authoritarian Dominance. *Personality and Social Psychology Bulletin, 30*(7), 824-837. doi:10.1177/0146167204264333
- Van Roon, A. M., Mulder, L. J. M., Althaus, M., & Mulder, G. (2004). Introducing a baroreflex model for studying cardiovascular effects of mental workload. *Psychophysiology, 41*(6), 961-981. doi:10.1111/j.1469-8986.2004.00251.x
- Vigil, J. M. (2010). Political leanings vary with facial expression processing and psychosocial. *Group Processes and Intergroup Relations, 13*(5), 547-558. doi:10.1177/1368430209356930gpir.sagepub.com
- Virkkunen, M., Rawlings, R., Tokola, R., Poland, R. E., Guidotti, A., Nemeroff, C., Bissette, G., Kalogeras, K., Karonen, S. L., & Linnoila, M. (1994). CSF biochemistries, glucose metabolism, and diurnal activity rhythms in alcoholic, violent offenders, fire setters, and healthy volunteers. *Archive of General Psychiatry, 51*(1), 20-27. doi:10.1001/archpsyc.1994.03950010020003
- Wager, D., Davidson, M. L., Hughes, B. L., Lindquist, M. A., & Ochsner, K., N. (2008). Prefrontal-Subcortical Pathways Mediating Successful Emotion Regulation. *Neuron, 59*(6), 1037-1050. doi:10.1016/j.neuron.2008.09.006

- Waldvogel, D., Gelderen, P. van, Muellbacher, W., Ziemann, U., Immisch, I., & Hallett, M. (2000). The relative metabolic demand of inhibition and excitation. *Nature*, *406*(6799), 995-998. doi:10.1038/35023171
- Waller, J. (2002). *Becoming Evil: How Ordinary People Commit Genocide and Mass Killing*. Oxford: Oxford University Press.
- Watson, D., Clark, L.A., & Tellegen, A., (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070. doi:10.1037/0022-3514.54.6.1063
- Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, *138*(4), 775-808. doi:10.1037/a0027600
- Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, *77*(4), 273-295. doi:10.1037/h0032351
- Weiss, D. J. (2001). Deception by researchers is necessary and not necessarily evil. *Behavioral and Brain Sciences*, *24*(3), 431-432.
- Weissflog, M. J., van Noordt, S. J. R., Choma, B. L., Dywan, J., & Segalowitz, S. J. (2010). Sociopolitical ideology and electrocortical responses. *Psychophysiology*, *47*(Sup), S24.
- Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*(1), 411-418.
- Wheat, A. L., & Larkin, K. T. (2010). Biofeedback of heart rate variability and related physiology: a critical review. *Applied Psychophysiology Biofeedback*, *35*(3), 229-242. doi:10.1007/s10484-010-9133-y
- Wiener, R. L., & Erker, P. V. (1986). The effects of prebriefing misinformed research participants on their attributions of responsibility. *Journal of Psychology*, *120*(4), 397-410. doi:10.1080/00223980.1986.9712638
- Williams, G. C. (1988). Reply to Comments on "Huxley's Evolution and Ethics in Sociobiological Perspective." *Zygon*, *23*(4), 437-438. doi:10.1111/j.1467-9744.1988.tb00857.x
- Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehavioural Reviews*, *25*(4), 287-295. doi:10.1016/S0149-7634(01)00014-8

- Wilson, G. D. (1973). *The psychology of conservatism*. London: Academic Press.
- Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225-240. doi:10.1162/105474698565686
- Wright, R. (1994). *The Moral Animal: The New Science of Evolutionary Psychology*. New York: Pantheon.
- Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *The Journal of Neuroscience*, 29(26), 8525-8529. doi:10.1523/JNEUROSCI.2418-09.2009
- Yerkes, R. M., & Dodson, J. D. (1908). The Relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459-482. doi:10.1002/cne.920180503
- Young, P. T. (1936). *Motivation of Behavior*. New York: John Wiley.
- Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., & Damasio, A. (2010). Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. *Neuron*, 65(6), 845-851. doi:10.1016/j.neuron.2010.03.003
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151-175. doi:10.1037/0003-066X.35.2.151
- Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist*, 39(2), 117-123. doi:10.1037/0003-066X.39.2.117
- Zald, D. H., & Pardo, J. V. (1997). Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), 4119-4124. doi:10.1073/pnas.94.8.4119
- Zeigler-Hill, V., Southard, A. C., Archer, L. M., & Donohoe, P. L. (2013). Neuroticism and negative affect influence the reluctance to engage in destructive obedience in the Milgram paradigm. *Journal of Social Psychology*, 153(2), 161-174. doi:10.1080/00224545.2012.713041
- Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the study of children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives*, 6(2), 112-121. doi:10.1111/j.1750-8606.2011.0017

Annexe 1 – CERNI – 4/07/2013



CERNI
Comité d’Ethique pour les
Recherches Non Interventionnelles

Objet: Avis sur le projet « Soumission à l'autorité et idéologie : effets de l'orientation politique et de l'environnement idéologique sur le comportement d'obéissance destructrice » présenté par Laurent BEGUE du Laboratoire Interuniversitaire de Psychologie (UPMF) à Grenoble.

Compte tenu des éléments fournis dans la demande déposée par Laurent BEGUE, le Comité d’Ethique pour les Recherches Non Interventionnelles émet un **AVIS FAVORABLE**.

Numéro de l’avis : 2013-07-04-23

A Grenoble, le 4 juillet 2013

Le Président du CERNI,

Jean-Pierre Chevrot

Professeur, Université de Grenoble et Institut Universitaire de France

Annexe 2 – CPP – 13/02/2015

LM 328

**COMITE DE PROTECTION DES PERSONNES
Sud Est V**

Grenoble le, 10/02/2015

Président :
M. Jean GRUNWALD

Vice-Président :
Dr Daniel ANGLADE

Secrétaire Général :
Dr Dominique CHARLETY


Secrétaire Général Adjoint :
Dr Joël JUGE

Site internet
<http://www.cppsudest5.fr>

Secrétariat :
cppsudest5@chu-grenoble.fr

F. LUCZAK
N. CESTARO
Tel : 04 76 76 57 83
Fax : 04 76 76 51 77
GSM : 06 50 85 79 26

Domiciliation :

 **CHU de Grenoble**
Comité de Protection des
Personnes

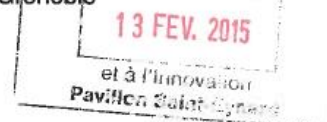
Adresse postale :
CS 10217
38043 GRENOBLE Cedex 9

Adresse géographique :
R.C. Haut
Hall Vercors
Bd de la Chantourne
38700 LA TRONCHE

N° SIRET : 130 016 017 00019
N°APE : 8412Z

IRB: 6705

Promoteur :
Madame H. SABBAH-GUILLAUME
Directrice de la Recherche Clinique et des
partenariats
Délégation à la Recherche Clinique
et à l'Innovation
Pavillon Saint-Synard
CHU de Grenoble



01. Réf. CPP: 15-CHUG-01
Ref Etude : PSM_EEG - 38RC14.420
N° ID RCB : 2014-A01823-44

Documents fournis :

Protocole : Version 2.0 du 15/12/2014 et Version 2.1 du 21/01/2015

Lettre d'information au patient : Version 2.0 du 15/12/2014 et Version 2.1 du 21/01/2015

Formulaire de consentement : Version 2.0 du 15/12/2014 et Version 2.1 du 21/01/2015

Promoteur : CHU de Grenoble
Investigateur Principal : Dr Laurent VERCUEIL - CHU de Grenoble
Type Recherche Biomédicale : AUTRE

Objet : **Avis favorable**

Madame,

Le Président du Comité, a donné un avis **FAVORABLE**, en date du **4 FEVRIER 2015**, pour le protocole reçu le 17/12/2014, ayant une demande de complément d'information le 14/01/2015, intitulé :

Etude des corrélats neurophysiologiques du lien entre la perception et l'action à travers divers protocoles d'études sensorimoteurs

après consultation des rapporteurs et examen par le Comité aux séances des 14/01/2015 et 04/02/2015.

Le promoteur est le CHU de GRENOBLE - DRCI - Pavillon Dauphiné - Hôpital La Tronche - CS 10217 - 38043 GRENOBLE Cedex 09

L'investigateur principal est le Dr Laurent VERCUEIL - Service d'explorations fonctionnelles du système nerveux (EFSN) - CHU de Grenoble.

N° RPPS : 10002435757

L'investigateur associé est le Pr Alexandre Krainik - Service de neuroradiologie - CHU de Grenoble
N° RPPS : 10001409167 / N° Conseil de l'ordre de l'Isère : 38/8323 / N° ADEL: 381083327

15-CHUG-01 / Avis -1/2

Les personnes ayant délibéré sur le projet sont :

1er collège :

1 - Recherche biomédicale

Titulaires : Dr ANGLADE Daniel

Dr BONAZ Bruno

Dr GRENIER Florian

Suppléants : Mme GOFTI-LAROCHE Leïla

Dr KUENTZ François

Dr SEIGNEURIN Arnaud

2 - Médecin généraliste :

Titulaire : Dr TIRARD Véronique

Suppléant : Dr FAURE Alain

3 - Pharmacien hospitalier

Titulaire : Mme CHARLETY Dominique

Suppléant : Mme DURAND Marjorie

4 - Infirmier : Néant :

Aucun membre délibérant du Comité n'est affecté par un conflit d'intérêt

2ème collège :

5 - Personnes qualifiées "éthique"

Suppléant : M RIBUOT Christophe

6 - Psychologue

Titulaire : M. GRUNWALD Jean

Suppléant : Mme NAEGELE Bernadette

7 - Travailleur social

Titulaire : Mme STEMPFLE Sandrine

8 - Personnes qualifiées "juridique"

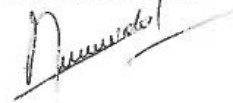
Titulaire : Mme DUCKI Myriam

9 - Représentants d'association agréée de malades et d'usagers du système de santé

Titulaire : Mme PELLOUX Colette

Je vous prie de croire, **Madame**, à l'expression de mes salutations distinguées.

Le Président,
Jean GRUNWALD



Annexe 3 – Notice d'informations études 3-6

**ETUDE SUR LES EFFETS DE LA DÉFIBRILLATION
SUR LE FONCTIONNEMENT COGNITIF**

NOTICE D'INFORMATION

.....

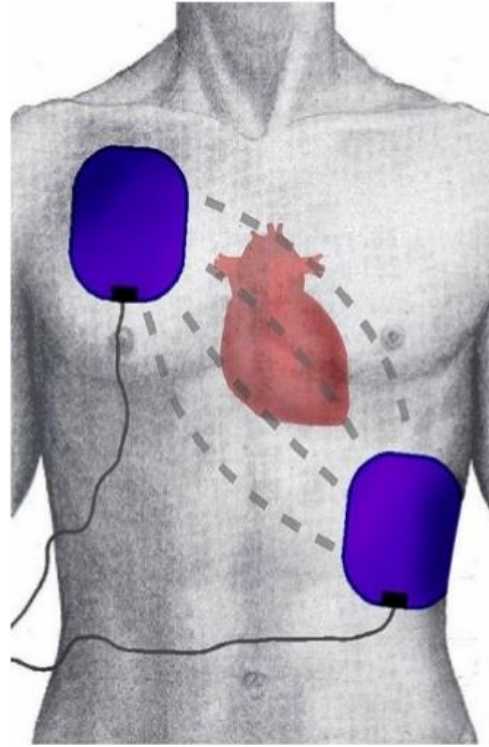
LABORATOIRE INTER-UNIVERSITAIRE DE PSYCHOLOGIE

The cover page features a horizontal strip of four images: an ECG waveform, a stethoscope, a brain scan, and a neuron.

CONTEXTE SCIENTIFIQUE

La défibrillation, appelée aussi choc électrique externe, est le geste médical consistant à faire passer volontairement et de manière brève un courant électrique dans le cœur lorsque celui-ci présente certains troubles du rythme appelés fibrillation, et destinés à rétablir un rythme cardiaque normal.

Souvent utilisée dans les services d'urgence, cette technique pourrait provoquer des troubles de la mémoire pendant quelques dizaines de minutes. Cet effet, non encore établi scientifiquement, pourrait être responsable d'importantes erreurs dans le diagnostic : la personne ayant subi une défibrillation pourrait se voir attribuer des troubles de la mémoire, alors que ces troubles ne seraient que temporaires. L'enjeu est donc d'améliorer le diagnostic des médecins auprès des personnes ayant subi une défibrillation.



OBJECTIFS DE L'ETUDE

L'objectif est de tester en laboratoire l'hypothèse d'un effet de la défibrillation sur le fonctionnement de la mémoire. Les sujets volontaires effectueront une tâche de mémoire et recevront des chocs électriques au cours de la procédure. L'intensité et la durée des chocs seront contrôlées de manière à reproduire les différents types de défibrillation pratiqués en milieu hospitalier.

METHODE

Communication par visio-conférence

Vous effectuerez cette expérience avec un(e) autre participant(e). Vous serez installé(e)s dans deux salles différentes, et un système de visio-conférence vous permettra de voir et d'entendre l'autre participant(e) (mais cette personne ne pourra ni vous voir, ni vous entendre).

.....
Deux catégories de participant(e)s ont été recrutées. Vous appartenez à la catégorie 2 et l'autre participant(e) appartient à la catégorie 1

Sujets Catégorie 1

La catégorie 1 est constituée de personnes ayant accepté de recevoir des chocs électriques. Des précautions ont été prises pour diminuer la dangerosité de la démarche :

- (i) un médecin sera présent dans la salle où seront installées les personnes recevant les chocs,
- (ii) les chocs ne seront pas envoyés au niveau du thorax mais au niveau des membres supérieurs (bras),
- (iii) les personnes seront attachées à leur chaise de manière à ce que leurs mouvements ne les fassent pas tomber.

Sujets Catégorie 2

La Catégorie 2 est constituée de personnes chargées d'envoyer les chocs électriques externes à l'autre participant(e).

Impartialité de la procédure et fiabilité des résultats :

Le recours à des exécutants externes indépendants vis-à-vis de l'expérience nous permet de garantir la validité de nos résultats. Ce point est particulièrement important et c'est pourquoi vous êtes en charge de l'envoi des chocs (et non l'expérimentateur).

TEST DE MEMOIRE

Le sujet Catégorie 1 effectuera un test de mémoire et le sujet Catégorie 2 aura à envoyer des chocs électriques pendant ce test de mémoire

Sujet Catégorie 1

Le test de mémoire est constitué de deux parties : mémorisation et rappel. Le sujet Catégorie 1 apprendra une liste de paires de mots avant l'expérience (ex : ciel – nuageux ; voiture – ancienne ; etc.) et effectuera un test de rappel pendant l'expérience. Pendant l'expérience, le médecin donnera à chaque essai le premier terme d'une paire de mots (ex : ciel) puis une liste de 4 mots possibles (ex : bleu, sombre, nuageux, éclairé). Le sujet Catégorie 1 aura à retrouver le deuxième terme (ici : nuageux).

Sujet Catégorie 2

Les sujets catégorie 2 auront à suivre les consignes apparaissant à l'écran tout au long de l'expérience. De manière générale, **il vous sera demandé** :

- (i) d'appuyer sur le **bouton bleu** du clavier pour passer à la question suivante
- (ii) d'appuyer sur le **bouton rouge** du clavier pour envoyer le choc électrique.

Pour les essais prévoyant l'envoi d'un choc électrique, une consigne vous demandera d'appuyer sur le **bouton rouge** avant la question. Il vous sera ensuite demandé d'appuyer sur le **bouton bleu**.

Pour les essais ne prévoyant pas l'envoi d'un choc électrique, une consigne vous demandera d'appuyer directement sur le **bouton bleu** pour passer à la question suivante.

DEROULEMENT DE L'EXPERIENCE

Test de rappel en 6 parties

Le test de rappel comprend 6 parties (avec 5 questions par partie). La figure ci-dessous illustre le déroulement de l'expérience. Dans les parties "sans chocs", le médecin indiquera après chaque réponse si celle-ci est correcte ou non. Dans les parties avec chocs, le médecin ne donnera aucun retour et un choc sera systématiquement envoyé avant chaque question.

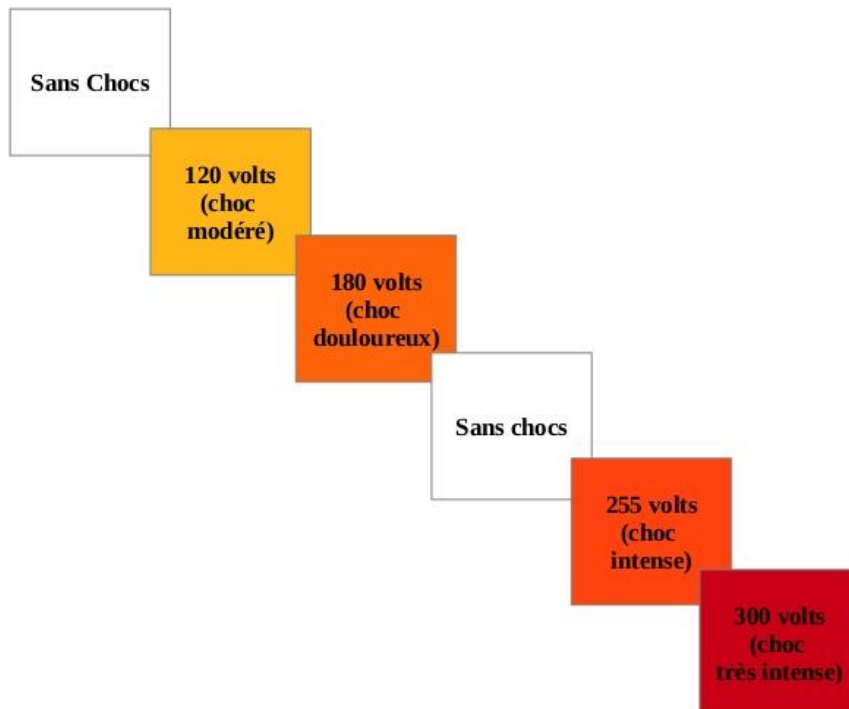


FIGURE 1 | Déroulement de l'expérience en 6 parties. Chaque partie comprend plusieurs questions. L'intensité des chocs augmentera au cours de la procédure.

Annexe 4 – Dernière page de la Notice d'informations étude 6



DEROULEMENT DE L'EXPERIENCE

Test de rappel en 9 parties

Le test de rappel comprend 9 parties (avec 7 questions par partie). La figure ci-dessous illustre le déroulement de l'expérience. Dans les parties "sans chocs", le médecin indiquera après chaque réponse si celle-ci est correcte ou non. Dans les parties avec chocs, le médecin ne donnera aucun retour et un choc sera systématiquement envoyé avant chaque question.

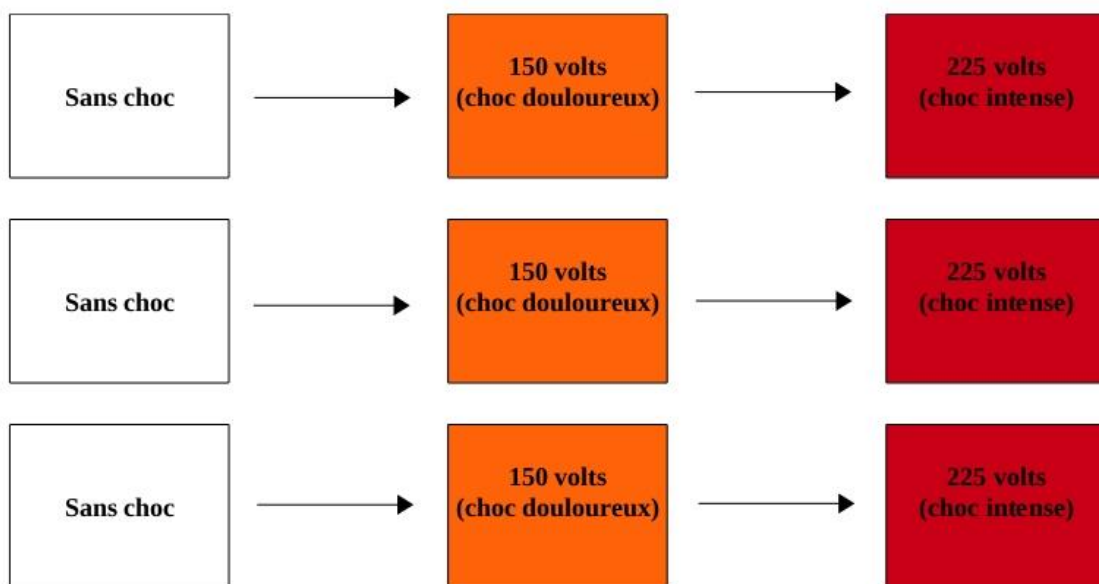


FIGURE 1 | Déroulement de l'expérience en 9 parties. Chaque partie comprend plusieurs questions. L'intensité des chocs augmentera au cours de la procédure.

*Annexe 5 – Procédure de recrutement études 3-6***Recrutement des participants aux expériences de soumission à l'autorité**

Pour le recrutement et l'inclusion des participants aux études 3-6, nous avons systématiquement suivi la procédure décrite ci-dessous.

1. Recrutement

Les participants étaient informés des études en cours via différentes modalités de communication : flyers et affiches (*cf.*, annexe 7), annonces dans la presse régionale, annonce sur la plateforme du RISC (Relai d'Information sur le Sciences Cognitives), sollicitation directe par mail.

2. Inclusion

Les personnes volontaires devaient se rendre sur un blog proposant une présentation plus extensive des études en cours (*cf.* annexe 8). Il était demandé aux personnes intéressées de vérifier qu'elles remplissaient bien les conditions suivantes :

- être âgé de plus de 18 ans le jour de l'expérience
- être âgé de moins de 50 ans le jour de l'expérience
- ne pas présenter de trouble(s) neurologique(s)
- ne pas présenter de trouble(s) psychiatriques(s)
- ne pas présenter de trouble(s) hormonal-naux
- ne pas présenter de trouble(s) cardio-vasculaire(s)
- ne pas prendre de traitement(s) médicamenteux régulier(s) (excepté contraception ; en cas de doute sur ce point, envoyez-nous un mail)
- avoir une vue normale ou corrigée
- avoir une audition normale ou corrigée
- être de langue maternelle française

En cas de satisfaction de ces critères, les personnes volontaires devaient renseigner un formulaire en ligne accessible à partir du blog (*cf.*, annexe 5). Les personnes satisfaisant les critères d'inclusion (e.g., absence de pathologie cardiaque, non connaissance de Milgram) étaient recontactées par mail et invitées à s'inscrire sur un planning en ligne.

Annexe 6 – Formulaire d'inclusion en ligne études 2-6

1. Veuillez indiquer votre numéro de téléphone ci-dessous *

2. Veuillez indiquer votre adresse mail ci-dessous *

3. Veuillez indiquer votre age ci-dessous *

4. Langue maternelle française *

Mark only one oval.

Non

Oui

5. Indication de la présence d'un trouble neurologique et/ou psychiatrique *

Mark only one oval.

Oui

Non

6. Indication de la présence d'un trouble cardio-vasculaire *

Mark only one oval.

Non

Oui

7. Indication de la présence d'un trouble endocrinien ou immunologique *

Mark only one oval.

Oui

Non

8. Indication de la présence d'un trouble auditif non corrigé *

Mark only one oval.

Non

Oui

9. **Indication de la présence d'un trouble visuel non corrigé ***

Mark only one oval.

- Oui
 Non

10. **Vous êtes sous traitement médicamenteux régulier (autre que contraception) ***

Mark only one oval.

- Non
 Oui

Eléments de culture académique

La dernière partie de ce formulaire porte sur vos connaissances académiques

11. **Connaissez-vous cet auteur et ses travaux : Shérif (1935) ***

Mark only one oval.

- Oui
 Non

12. **Connaissez-vous cet auteur et ses travaux : Asch (1951) ***

Mark only one oval.

- Non
 Oui

13. **Connaissez-vous cet auteur et ses travaux : Moscovici (1969) ***

Mark only one oval.

- Oui
 Non

14. **Connaissez-vous cet auteur et ses travaux : Milgram (1974) ***

Mark only one oval.

- Non
 Oui

15. **Connaissez-vous cet auteur et ses travaux : Allport (1954) ***

Mark only one oval.

- Oui
 Non

16. **Connaissez-vous cet auteur et ses travaux : Festinger (1957) ***

Mark only one oval.

- Non
 Oui

17. **Connaissez-vous cet auteur et ses travaux : Beauvois (2002) ***

Mark only one oval.

Oui

Non

18. **Avez-vous vu l'émission « Le jeu de la mort » ? ***

Mark only one oval.

Non

Oui

19. **Avez-vous vu l'émission "Le cerveau et ses automatismes" ? ***

Mark only one oval.

Oui

Non

Annexe 7 – Formulaire d'inclusion études 5 et 6

QUESTIONNAIRE D'INCLUSION VOLONTAIRE ADULTE**Critères d'inclusion**

- Age entre 18 et 70 ans	oui	non
- Niveau d'étude Bac minimum	oui	non
- Vision normale ou corrigée à la normale	oui	non
- Examen médical préalable de pré-inclusion	oui	non
- Affiliation à ou bénéficiaire d'un régime de sécurité sociale	oui	non
- Consentement éclairé signé	oui	non

Inclus si six réponses oui

Critères de non inclusion :

- Sujets inclus dans une autre expérimentation clinique et/ou thérapeutique en cours	oui	non
- Troubles important de l'audition et de la vision	oui	non
- Trouble du langage (aphasie, dysphasie, dysarthrie, bégaiement, etc.)	oui	non
- Pathologies neurologiques ou neuropsychiatriques passées ou actuelles	oui	non
- Traitement médicamenteux susceptible de moduler l'activité cérébrale : benzodiazépine, antidépresseurs, neuroleptiques, lithium, etc.	oui	non
- Contre indications à l'IRM (cf. questionnaire au verso)	oui	non
- Femme enceinte		
Test urinaire de grossesse proposé – G test positif	oui	non NA
Si test refusé par volontaire – signature formulaire de décharge (Annexe B)		
- Femme parturiente	oui	non NA
- Femme allaitant	oui	non NA
- Sujet de toutes catégories de personnes protégées	oui	non

Si une réponse oui, exclusion

Critères d'exclusion :

- Ingestion d'alcool	oui	non
----------------------	-----	-----

Annexe 8 – Exemple de Flyer utilisé pour le recrutement

40€ pour 60 minutes de votre temps

Participez aux expériences de recherche en
Psychophysiology



experience.screen@gmail.com

Nous nous intéressons au **fonctionnement cardiaque** et à la **mémoire humaine**.

Les études proposées sont « non invasives » et anonymes : aucune injection, aucune prise de médicament. Les mesures physiologiques réalisées sont effectuées à l'aide de capteurs placés à la surface de la peau.

Exemple d'expériences : entraînement à la régulation de la fréquence cardiaque à l'aide d'un logiciel dédié, repérer et sélectionner des images sur un écran d'ordinateur, etc.

Qui peut participer ?

Toute personne majeure, et dont le français est la langue maternelle.



Comment participer ?

Il vous suffit d'envoyer un mail à l'adresse mail indiquée ci-dessus. Vous recevrez en réponse un mail d'information et une proposition d'inscription.

Lieu :

MSH-Alpes
Domaine universitaire de Saint-Martin-d'Hères
1221 Avenue Centrale

Annexe 9 - Page de Blog pour étude 4



Accueil	ETUDE 1	ETUDE 2	S'inscrire à une étude	Plan d'accès	
-------------------------	-------------------------	-------------------------	--	------------------------------	--

ETUDE 1

Titre de l'étude
Fonction cardiaque et Mémoire

Chercheur titulaire responsable scientifique du projet
Pr. Laurent Bègue, MSH-Alpes, Grenoble

Lieu de recherche
SCREEN, MSH-Alpes, Campus Universitaire, Grenoble

Dates
Du 17/02/15 au 24/04/15 (du lundi au vendredi inclus)

Durée totale de l'expérience
60mn en moyenne (explication + installation + passation)

Indemnité
20 euros

Objectifs
Evaluer le rôle de certains facteurs pouvant influencer l'activité cardiaque ainsi que la mémoire humaine.

Les tâches que vous aurez à effectuer
(i) Réaliser une tâche consistant à repérer et sélectionner des éléments visuels. Votre activité cardiaque sera enregistrée pendant cet exercice (électrocardiogramme).
(ii) Réaliser une tâche de mémoire en collaboration avec un(e) autre participant(e).

Risques possibles
Cette recherche n'implique aucun risque ou inconfort autres que ceux de la vie quotidienne. L'information électrique venant de votre corps est enregistrée à l'aide d'électrodes, mais rien ne sera envoyé par les appareils utilisés.

Posez des questions à tout moment
Vous pouvez poser des questions à propos de la recherche à tout moment en communiquant avec le responsable scientifique du projet par courrier électronique à : experience.screen@gmail.com

Contact
experience.screen@gmail.com

Laboratoires impliqués
Laboratoire Inter-universitaire de Psychologie (Université Grenoble-Alpes)
Site web : <http://www.liip.univ-savoie.fr/>
Laboratoire de Psychologie et de Neurocognition (CNRS, Université Grenoble-Alpes)
Site web : <http://webcom.upmf-grenoble.fr/LPNC/>

LISTE DES ABREVIATIONS

AMY	Amygdale
BA	Brodmann area
CCA	Cortex cingulaire antérieur
COF	Cortex orbitofrontal
CPF	Cortex préfrontal
CPFDL	Cortex préfrontal dorsolatéral
CPFVM	Cortex préfrontal ventromédian
CS	Corrugator supercillii
EC	Effortful control
ECG	Électrocardiogramme
EEG	Électroencéphalogramme
EFE	Expression faciale émotionnelle
EMG	Électromyographie
EVI	Environnement virtuel immersif
fNIRS	Functional near-infrared spectroscopy
Hb	Hémoglobine
HF-HRV	High frequencies heart rate variability
HPA	Hypothalamic-pituitary-adrenal
HRV	Heart rate variability
HRVb	HRV basale
IFG	Inferior frontal gyrus
IRI	Indice de réactivité interpersonnelle
IRMf	Imagerie par résonance magnétique fonctionnelle
ISRS	Inhibiteur sélectif de la recapture de la sérotonine
IVET	Immersive virtual environment technology
IVMOE	Immersive Video Milgram Obedience Experiment
LH-HRV	Low frequencies heart rate variability
MSC	Motivation à la sécurité collective
OD	Obéissance destructrice
PFC	Prefrontal cortex
RMSSD	Root mean squared successive differences
RSA	Respiratory sinus arrhythmia
RWA	Right-wing authoritarianism
SDO	Social dominance orientation
SNA	Système nerveux autonome
SNC	Système nerveux central
SNP	Système nerveux parasympathique
SNS	Système nerveux sympathique
VPL	Ventral posterolateral nucleus
ZM	Zygomaticus major

LISTE DES TABLEAUX

Introduction	
1.	Prédictions du comportement adopté par soi dans la condition standard..... 42
2.	Résultats des expériences 1-4 reposant sur la manipulation de la variable « proximité de la victime » (Milgram, 1974)..... 46
3.	Réplications interculturelles des résultats de Milgram (condition standard)..... 48
4.	Description des 23 expériences réalisées par Milgram..... 50
5.	Résultats de la méta-analyse de Haslam et al. (2014)..... 54
6.	Protocole expérimental et principales mesures des études relatives aux déterminants dispositionnels de l'OD..... 58
7.	Compétences socio-émotionnelles associées à l'empathie selon Decety et Jackson (2004) 97
8.	Stades phylogénétiques du contrôle neural du système nerveux autonome selon Porges (2007)..... 112
9.	Taux de suspicion dans les expériences de Milgram (Milgram, 1974)..... 127
10.	Taux d'obéissance dans la condition standard de Milgram (1963) et dans l'étude d'O'Leary et al. (1970) reposant sur le jeu de rôle actif..... 130
11.	Résultats de l'étude de Dambrun et Vatiné (2010)..... 140
Manuscrit 1	
1.	Descriptive statistics and correlations between authoritarian attitudes, dispositional altruism and moral outcomes..... 164
2.	Multiple regression examining the associations between ideological attitudes and responses to moral dilemmas..... 165
Manuscrit 2	
1.	Descriptive statistics and correlations between heart rate variability parameters and authoritarian attitudes..... 196
2.	Linear regressions examining the associations between heart rate variability parameters and authoritarian attitudes..... 197
Manuscrit 3	
1.	Distribution of break-off points..... 237
2.	Descriptive statistics and correlations between obedience, right-wing authoritarianism and physiological parameters..... 238
3.	Regression analyses examining the associations between physiological parameters and obedience..... 240
Manuscrit 4	
1.	Distribution of break-off points..... 284
2.	Relative obedience as a function of self-regulation condition and RWA scores (median split)..... 291
3.	Multiple regression examining the associations between relative obedience and the expected predicting variables (resting HRV and RWA)..... 306
4.	Multiple regression examining the behavioral consequences of SRE..... 307
Manuscrit 5	
1.	Distribution of break-off points..... 328
Manuscrit 6	
1.	Binary logistic regression examining the association between oxy-Hb levels (VMPFC, DLPFC) and absolute obedience..... 368
2.	Multiple regression examining the association between oxy-Hb levels (VMPFC, DLPFC) and relative obedience..... 369
3.	Binary logistic regression examining the association between oxy-Hb levels (right OFC, left OFC) and absolute obedience..... 370
4.	Multiple regression examining the association between oxy-Hb levels (right OFC, left OFC) and relative obedience..... 370

LISTE DES FIGURES

Introduction	
1. Disposition générale de l'espace expérimental. (Milgram, 1974).....	38
2. Représentation du générateur de chocs électriques. (Milgram, 1974).....	39
3. Prédications comportementales par un échantillon de psychiatres. (Milgram, 1974).....	43
4. Pourcentages d'obéissance dans les expériences 1-4. (Milgram, 1974).....	45
5. Représentation schématique des travaux de Milgram. (Haslam et al., 2014).....	53
6. Modèle duel motivationnel. (Duckitt & Sibley, 2010).....	69
7. Modèle de la sécurité collective des attitudes autoritaires. (Jugert & Duckitt, 2009).....	71
8. Mécanismes par lesquels l'auto-sanction morale est sélectivement engagée ou désengagée à différents points du processus d'autorégulation. (Bandura, 1986).....	79
9. Structures cérébrales impliquées dans le jugement moral.....	89
10. Représentation schématique des principales composantes anatomiques de la neuromatrice de la douleur. (Decety & Lamm, 2009).....	93
11. Distinction des composantes sensoridiscriminantes et affectives-motivationnelles de la neuromatrice de la douleur.....	94
12. Décours temporel d'un essai avec choc électrique. (Cheetham et al., 2009).....	99
13. Réponses hémodynamiques significatives pendant l'observation de la douleur de l'avatar. (Cheetham et al., 2009).....	101
14. Modèle de la sensibilité émotionnelle vs régulation émotionnelle. (Koole, 2009).....	103
15. Illustration de la loi Yerkes-Dodson (1908).....	105
16. Le complexe QRS réfère à 3 ondes accolées correspondant à la dépolarisation des ventricules.....	114
17. Contrôle neural du cœur.....	119
18. Scénario expérimental. (Slater et al., 2006).....	134
19. Interface de l'« Immersive Video Milgram Obedience Experiment » (IVMOE). (Dambrun & Vatiné, 2010).....	137
Manuscrit 1	
1. Standardized regression coefficients (a) for the relationship between SDO and Judgment (Trolley dilemma) as mediated by altruism, (b) for the relationship between SDO and Judgment (Footbridge dilemma) as mediated by altruism, (c) for the relationship between SDO and Choice (Footbridge dilemma) as mediated by altruism.....	168
Manuscrit 2	
1. Scatter plots for the association between heart rate variability parameters and authoritarian attitudes.....	198
Manuscrit 3	
1. Experimental procedure.....	234
2. Scatter plots, densities, and Pearson correlation coefficients between recorded variables.....	239
3. Heart rate variability (RMSSD) and heart rate (beats per minute) changes among obedient and disobedient participants.....	242
4. Standardized regression coefficients (a) for the relationship between RMSSD (at rest) and relative obedience as mediated by inhibitory control, (b) for the relationship between HF-HRV (at rest) and relative obedience as mediated by inhibitory control (c) for the relationship between VR and relative obedience as mediated by inhibitory control, (d) for the relationship between right-wing authoritarianism and relative obedience as mediated by inhibitory control.....	245
Manuscrit 4	
1. Experimental procedure.....	281
2. Scatter plots, densities, and Pearson correlation coefficients between recorded variables (overall).....	285

3.	Scatter plots, densities, and Pearson correlation coefficients between recorded variables (low SRE condition).....	286
4.	Scatter plots, densities, and Pearson correlation coefficients between recorded variables (high SRE condition).....	287
5.	Heart rate variability (baseline 1) as a function of absolute obedience. Mean changes are shown with 95% CI.....	288
6.	Change in heart rate variability as a function of self-regulation condition (baseline 2 – baseline 1). Mean changes are shown with 95% CI.....	289
7.	Absolute obedience as a function of self-regulation condition and RWA scores (median split). Mean changes are shown with 95% CI.....	290
Manuscript 5		
1.	Experimental procedure.....	324
2.	Contrast topographies (difference of Z scores) in the theta band power between the disobedient and obedient groups across time.....	330
Manuscript 6		
1.	Overview of the experimental procedure.....	359
2.	Overview of the timeline of a pain trial.....	360
3.	Channels template. The 48-channel NIRS system consisted in 8 emitters (yellow circles) and 8 receivers (blue circles) for each hemisphere, resulting in 24-left and 24-right measurement points (yellow lengthened shapes).....	362
4.	Channels for the orbitofrontal-ventromedial prefrontal cortex (green) and for the dorsolateral prefrontal cortex (purple). Channels for the orbitofrontal cortex are outlined in red.....	365
5.	(A) Oxy-Hb levels among obedient and disobedient participants for condition 1 (No Pain). (B) Oxy-Hb levels among obedient and disobedient participants for condition 2 (the two first “150V” and “225V” blocks).....	367
Discussion		
20.	Zygomaticus major et corrugator supercilii.....	403