



L'influence persistante des fausses informations malgré leur correction : étude de validité et de nouveaux déterminants

Victor LAURENT

Thèse de doctorat en vue de l'obtention du titre de Docteur en psychologie,
Sous la direction de Stéphane RUSINEK (Directeur de thèse) & Thierry KOSINSKI (Co-encadrant)

Réalisée à l'Université de Lille, au sein du laboratoire PSITEC et de l'École Doctorale
des Sciences de l'Homme et de la Société,
Présentée et soutenue publiquement le 13 novembre 2024.

Membres du jury :

BERJOT Sophie,	Présidente du jury et examinatrice, Professeure des universités, Université de Reims Champagne-Ardenne.
DELOUVEE Sylvain,	Rapporteur, Maître de conférences – HDR, Université Rennes 2.
FOINTIAT Valérie,	Rapporteuse, Professeure des universités, Aix-Marseille Université.
RUSINEK Stéphane,	Directeur de thèse, Professeur des universités, Université de Lille.
KOSINSKI Thierry.	Co-encadrant de thèse, Maître de conférences, Université de Lille.



Remerciements

Je souhaite avant tout remercier le docteur Thierry Kosinski, d'avoir été toujours si disponible, réactif, encourageant et de si bons conseils. Merci également d'être, depuis plusieurs années maintenant, un modèle de rigueur scientifique et d'exigence qui aura profondément inspiré et marqué ma manière d'appréhender la recherche. Je remercie également le professeur Stéphane Rusinek de m'avoir accordé sa confiance en acceptant de diriger cette thèse tout en m'offrant une autonomie nécessaire à mon épanouissement.

Je tiens aussi à adresser mes plus sincères remerciements aux membres du jury, les professeurs Sophie Berjot et Valérie Fointiat ainsi que le docteur Sylvain Delouée, d'avoir bien voulu mettre à contribution leur grande expertise dans l'évaluation de cette thèse. C'est un honneur immense et j'en suis extrêmement reconnaissant.

Plusieurs personnes m'ont aidé dans la réalisation de ce travail. Je pense au professeur Claire Hofer dans l'aide à la traduction d'un texte, au docteur et capitaine Frédérique Gros pour nos discussions lors de la genèse de ce projet et sa relecture des textes militaires, à Solidarité SIDA, Tom Barris, Bill Posters et Daniel Howe d'avoir accepté que leur *deepfake* soit utilisé dans ce travail, à mon ami Damien Billion pour son concours dans l'édition d'un *deepfake*, et à tous les étudiants de licence et master en psychologie qui ont apporté leurs réflexions et contribué à la récolte de données.

Au cours de ce doctorat, j'ai la chance d'avoir bénéficié d'un environnement professionnel stimulant et agréable. A mes collègues du service de recherche en psychologie de la Marine nationale, Rémi, Joanna, Stéphane, Mickaël, René-Paul, Thierry, Sabine, Delphine, Emilie ; à mes collègues doctorants du laboratoire PSITEC, Andres, Antoine, Marion, Alexandra, Julie, Florent, Elliot et les autres. Merci à vous d'avoir fait de ce travail solitaire une belle aventure partagée.

J'ai aussi l'avantage d'avoir été entouré par des proches d'une qualité infinie. Merci à tous mes amis ainsi qu'à Maxime, Ornella et mes parents, pour leur présence réconfortante et de n'avoir jamais douté de moi.

Enfin, je souhaite exprimer ma plus grande gratitude à Grâce. J'ai passé trois années de doctorat aussi enrichissantes qu'intenses, occupant beaucoup de mes vacances et week-ends à la lecture d'articles, à de la rédaction, ou à simplement vaguer dans les pensées brumeuses d'un doctorant. Ces trois années sont pourtant restées merveilleuses grâce à elle. Merci de m'avoir soutenu, compris, encouragé, accompagné pendant tout ce temps et vivement ce week-end !

A tous, merci pour l'influence persistante que vous aurez sur moi et sur la représentation positive que je conserverai de cette aventure.

Table des matières

AVANT-PROPOS	X
---------------------	----------

CHAPITRE 1. LES FAUSSES INFORMATIONS, UNE PROBLEMATIQUE CONTEMPORAINE	1
--	----------

1.1 DEFINITION DES FAUSSES INFORMATIONS	2
--	----------

1.1.1 LA NOTION D'INFORMATION	2
-------------------------------	---

1.1.2 LA NOTION DE FAUX	3
-------------------------	---

1.1.3 LA NOTION DE FAUSSES INFORMATIONS	4
---	---

1.2 IMPORTANCE DE LA PROBLEMATIQUE DES FAUSSES INFORMATIONS	6
--	----------

1.2.1 DIFFUSION DES FAUSSES INFORMATIONS	6
--	---

1.2.2 ADHESION AUX FAUSSES INFORMATIONS	9
---	---

1.3 UBIQUITE DES CONSEQUENCES DES FAUSSES INFORMATIONS	13
---	-----------

1.4 APPROCHE CORRECTIVE FACE AUX FAUSSES INFORMATIONS	18
--	-----------

CHAPITRE 2. LES FAUSSES INFORMATIONS PERSISTENT MALGRE LEUR CORRECTION	20
---	-----------

2.1 CONCEPTUALISATION CLASSIQUE DE L'EFFET D'INFLUENCE CONTINUE	21
--	-----------

2.1.1 DEFINITION	21
------------------	----

2.1.2 HISTORIQUE ET PARADIGME CLASSIQUE	23
---	----

2.2 EXPLICATIONS THEORIQUES	25
------------------------------------	-----------

2.2.1 APPROCHE COGNITIVE	25
--------------------------	----

2.2.2 APPROCHE MOTIVATIONNELLE	33
--------------------------------	----

2.2.3 APPROCHE RATIONNELLE	37
----------------------------	----

2.3 AUTRES POSSIBILITES A LA SUITE DE LA CORRECTION D'UNE FAUSSE INFORMATION	40
---	-----------

2.3.1 EFFET BOOMERANG	40
-----------------------	----

2.3.2 SUR-CORRECTION	49
----------------------	----

2.3.3	RETOUR OU MAINTIEN DE LA LIGNE DE BASE	55
CHAPITRE3. REPLICABILITE ET VALIDITE DE L'EFFET D'INFLUENCE CONTINUE		60
<hr/>		
3.1	REPLICABILITE CONCEPTUELLE DE L'EFFET D'INFLUENCE CONTINUE	62
3.1.1	VARIATIONS DANS LES PROCEDURES	62
3.1.2	VARIATIONS DANS LES MESURES	63
3.1.3	VARIATIONS DANS LES DOMAINES D'APPLICATION	66
3.2	EXPERIENCE 1. REPLICABILITE CONCEPTUELLE AU DOMAINE MILITAIRE	70
3.2.1	METHODE	70
3.2.2	RESULTATS	75
3.2.3	DISCUSSION	77
3.3	EXPERIENCE 2. REPLICABILITE PROCHE AVEC UNE POPULATION FRANCOPHONE	82
3.3.1	METHODE	86
3.3.2	RESULTATS	89
3.3.3	DISCUSSION	90
CHAPITRE4. DETERMINANTS DE L'EFFET D'INFLUENCE CONTINUE		93
<hr/>		
4.1	VARIABLES EN LIEN AVEC LA FAUSSE INFORMATION	95
4.1.1	CARACTERE EXPLICITE DE LA FAUSSE INFORMATION	95
4.1.2	CREDIBILITE DE LA FAUSSE INFORMATION	97
4.1.3	VALENCE DE LA FAUSSE INFORMATION	97
4.1.4	CENTRALITE DE LA FAUSSE INFORMATION	99
4.1.5	REPETITION DE LA FAUSSE INFORMATION	101
4.2	EXPERIENCE 3. IMPACT DU FORMAT DE LA FAUSSE INFORMATION SUR L'EFFET D'INFLUENCE CONTINUE	102
4.2.1	INTRODUCTION	102

4.2.2	METHODOLOGIE	107
4.2.3	RESULTATS	115
4.2.4	DISCUSSION	117
4.3	VARIABLES EN LIEN AVEC LA CORRECTION	123
4.3.1	FORMULATION DE LA CORRECTION	123
4.3.2	FORMAT DE LA CORRECTION	127
4.3.3	CREDIBILITE DE LA CORRECTION	130
4.3.4	REPETITION DE LA CORRECTION	132
4.3.5	COHERENCE ENTRE LA CORRECTION ET LES ATTITUDES INITIALES	133
4.4	VARIABLES EN LIEN AVEC LA SITUATION	137
4.4.1	EFFET DU TEMPS	137
4.4.2	CHARGE COGNITIVE	139
4.4.3	ETAT EMOTIONNEL	139
4.5	EXPERIENCE 4. IMPACT DE L'IMPLICATION SUR L'EFFET D'INFLUENCE CONTINUE	142
4.5.1	INTRODUCTION	142
4.5.2	METHODOLOGIE	144
4.5.3	RESULTATS	149
4.5.4	DISCUSSION	151
4.6	EXPERIENCE 5. IMPACT DE VARIABLES DISPOSITIONNELLES SUR L'EFFET D'INFLUENCE CONTINUE	156
4.6.1	INTRODUCTION	156
4.6.2	METHODOLOGIE	162
4.6.3	RESULTATS	166
4.6.4	DISCUSSION	168
CHAPITRE 5. CONCLUSION		173
5.1	APPORTS	175

	viii
5.1.1 APPORTS THEORIQUES	175
5.1.2 APPORTS PRATIQUES	178
5.2 PERSPECTIVES	180
5.2.1 RECHERCHES FUTURES	180
5.2.2 LIMITE COMMUNE AUX EXPERIENCES	186
5.2.3 LIMITES ET ALTERNATIVES A L'APPROCHE CORRECTIVE	186
REFERENCES	192

Avant-propos

Les Fausses Informations (FI) sont depuis longtemps des faits notables de société (Basso & Kanellos, 2019 ; Watson, 2018). A l'époque de Cléopâtre, de fausses rumeurs concernant sa vie privée auraient été diffusées par ses adversaires politiques (Basso & Kanellos, 2019 ; Bièvre-Perrin, 2018). A la même période, Octave aurait diffusé des FI contre son rival Marc-Antoine en inscrivant des affirmations diffamatoires sur des pièces de monnaie (Posetti & Matthews, 2018). Alexandre le Grand aurait quant à lui laissé des armures taillées pour des hommes fictifs de plus de deux mètres sur les traces de ses campements. Dans la même idée, afin d'octroyer à son armée l'apparence d'un effectif plus grand qu'il ne l'était, Gengis Khan aurait fait porter plusieurs torches à chacun de ses soldats et attacher de larges objets derrière leurs chevaux (McKey, 2014 ; Rouse, s. d.) (voir aussi Basso & Kanellos, 2019 ; Posetti & Matthews, 2018).

Ces quelques cas de désinformation illustrent que les FI ne sont pas un phénomène récent (voir aussi Basso & Kanellos, 2019 ; Posetti & Matthews, 2018). Elles suscitent toutefois aujourd'hui des préoccupations majeures. Fin 2018 en France, une loi pour lutter contre les fausses informations s'ajoutait au dispositif législatif déjà présent. Quelques temps plus tard, le 29 septembre 2021, le président de la République signait une lettre demandant au sociologue Gérard Bronner de constituer une commission autour du sujet des (fausses) informations à l'ère du numérique (Bronner et al., 2022). Moins d'un mois après, le 20 octobre 2021, la ministre des Armées de l'époque prononçait un discours (Parly, 2021) dans lequel elle reconnaissait officiellement la nécessité pour les armées françaises de lutter contre les attaques informationnelles. A l'international, des dispositions semblables sont prises, comme en Australie où le premier ministre a annoncé la mise en place d'une législation spécifique pour lutter contre les tentatives étrangères de tromperie (Turnbull, 2017). Enfin, cette dynamique est aussi observable sur le plan scientifique, le sujet des fausses informations générant chaque

année plus de recherches sur Google Scholar (Greene et al., 2023). Le présent travail s'insère dans cette tendance en s'intéressant à l'influence des FI. Une première partie introductive s'attachera à étudier l'importance de leur diffusion, de leur adhésion et de leurs conséquences aujourd'hui.

Une stratégie fréquemment employée pour lutter contre les FI consiste à les corriger. A leur échelle, plusieurs acteurs tels que des organisations civiles (Graves & Cherubini, 2016) ou politiques (e.g., NATO, 2020) se mobilisent pour apporter une correction aux faits erronés. Des travaux observent cependant que la correction d'une FI ne permettrait pas toujours de faire totalement disparaître son influence. Par exemple, une recherche observait que des participants qui apprenaient qu'un politicien avait été corrompu, avant d'apprendre que cette information était fautive, continuaient d'avoir une attitude plus négative envers l'homme politique par rapport à s'ils n'avaient pas reçu la FI (Guillory & Geraci, 2013). La littérature utilise la notion d'« effet d'influence continue » (EIC, Johnson & Seifert, 1994) pour définir cette persistance des FI malgré leur correction. L'objectif du second chapitre sera de définir cet effet et d'examiner les explications théoriques qui lui sont fréquemment données.

Il doit cependant être noté que l'observation d'un effet dans quelques travaux scientifiques n'implique pas que cet effet soit forcément valide. Si certaines recherches observent clairement l'EIC (Guillory & Geraci, 2013 ; Johnson & Seifert, 1994), d'autres suggèrent qu'il serait plutôt rare et ténu (Ecker & Rodricks, 2020). L'objectif du troisième chapitre sera alors d'étudier la validité de cet effet. Une analyse de la littérature sera menée puis complétée par deux expériences.

Une autre question importante est de déterminer dans quelles circonstances l'EIC est intense et dans lesquelles il est plus faible voire absent. Cette question est importante sur le plan théorique et pratique. Par exemple, observer que l'EIC est plus grand lorsque les ressources cognitives sont saturées permettrait d'inférer que des processus cognitifs le sous-tendent tout

en indiquant qu'une FI serait mieux corrigée si l'attention de la personne qui y adhère est stimulée. C'est pour ces raisons théoriques et pratiques que l'objectif du quatrième chapitre sera d'étudier des éléments qui modulent l'EIC. Cette analyse sera réalisée à partir de la littérature ainsi qu'avec la conduite de trois expériences. Ce travail se terminera en faisant une synthèse des éléments pratiques et théoriques discutés. L'ambition finale est d'offrir une meilleure compréhension de l'EIC, et d'optimiser ainsi la correction des informations fausses à une époque où elles prospèrent.

Note : Suivant les recommandations de Simmons et al. (2012), nous reportons, pour chacune des expériences menées dans ce travail, comment nous avons déterminé la taille d'échantillon, les exclusions de données (le cas échéant), toutes les manipulations et toutes les mesures utilisées. Par ailleurs, au sein de ce travail, des exemples d'informations ou d'items sont parfois donnés afin d'illustrer certaines expériences ou échelles. La plupart d'entre eux provenaient de contenus anglophones mais sont ici retranscrits en français. Sauf indication contraire, la plupart de ces traductions ont été réalisées de manière libre. Le contenu original ne sera pas indiqué dans le texte afin de simplifier sa lecture mais le lecteur intéressé pourra se référer à la source du contenu.

**Chapitre 1. Les fausses
informations, une
problématique contemporaine**

Après avoir défini comment seront entendues les FI dans ce travail, l'objectif de ce chapitre sera de mettre en exergue leur importance sociétale. Leur diffusion et l'adhésion qu'elles suscitent seront discutées puis leurs conséquences dans divers domaines seront mises en avant. Un aperçu de plusieurs actions déployées avec l'intention de les corriger sera ensuite exposé.

1.1 Définition des fausses informations

La notion de FI en appelle d'autres telles que celles de désinformation, de mésinformation, de fausses nouvelles ou encore de théories du complot. Au sein de cette partie, une définition générale des FI est d'abord proposée afin de les distinguer de ces autres termes. Cette définition est construite à partir d'une conceptualisation de la notion d'information et de la notion de faux. Il est important de souligner que l'ambition ici n'est pas de proposer une définition des FI pour la littérature mais simplement de mentionner comment sera comprise cette notion dans ce travail.

1.1.1 La notion d'information

Parmi les nombreuses manières de définir une information, une définition proposée par le Larousse est une « nouvelle communiquée par une agence de presse, un journal, la radio, la télévision » (Larousse, 2023). Un autre dictionnaire propose, entre autres, qu'une information est un « renseignement ou évènement qu'on porte à la connaissance d'une personne ou d'un public » (Le Robert, 2023). Dans une perspective plus générale, certains auteurs ont proposé qu'une information était une donnée (e.g., Buckland, 1991 ; Dougherty et al., 2010), soit un élément brut saisi en dehors de tout contexte (Ermine et al., 2012).

Dans le cadre de ce travail, la notion d'information ne limitera pas à celle provenant des médias. D'autres sources d'information seront admises, telles que des pairs, des experts ou des institutions (Savolainen & Kari, 2004). Elle ne se limitera pas non plus à l'idée de « nouvelle » ou de renseignement porté à la connaissance de quelqu'un. L'approche choisie au

sein de ce travail sera plus générale et considérera qu'un élément physique tel qu'un signal radar peut constituer une information. Enfin, plusieurs auteurs ont suggéré l'idée qu'une simple donnée (e.g., un point rond) ne serait pas suffisante pour être informative et ont ajouté que du contexte était nécessaire pour qu'elle le devienne (e.g., un point rond sur un écran radar) (Biermann et al., 2004 ; Chaudet, 2009 ; Floridi, 2010). En accord avec cette perspective, l'approche choisie au sein de ce travail sera de définir une information comme étant une « donnée contextualisée », c'est-à-dire une donnée à laquelle du sens est ajouté.

1.1.2 La notion de faux

Dans le cadre de ce travail, une première approche pour définir la véracité d'un élément sera de rejoindre la proposition de plusieurs auteurs qui est d'estimer qu'un élément est vrai s'il existe une relation de correspondance entre cet élément et l'état du monde (Ludwig, 2016 ; Revault d'Allonnes, 2018 ; Vraga & Bode, 2020). Cette idée de correspondance était déjà exprimée par Aristote à travers la formule « *Dire de ce qui est qu'il n'est pas, ou de ce qui n'est pas qu'il est, c'est le faux ; dire de ce qui est qu'il est, et de ce qui n'est pas qu'il n'est pas, c'est le vrai* » (Aristote, 1953, cité dans Ludwig, 2016 ; Haack, 2019). Concrètement, si une proposition affirme qu'il existe un changement climatique et qu'il existe effectivement un changement climatique, alors cette proposition serait vraie. De même, si une proposition indique qu'un local est vide alors qu'une personne située en face du même local voit qu'il est plein, alors cette proposition serait fausse.

Malgré son intérêt théorique, une limite à l'opérationnalisation de cette première définition est que l'état du monde n'est pas toujours connu. Par exemple, la plupart des individus ne sont pas en mesure de percevoir directement si le changement climatique est réel ou factice, s'il existe une relation de cause à effet entre un vaccin et l'autisme, ou s'il y avait bel et bien des armes de destruction massive en Irak en 2003 lorsque les Etats-Unis ont envahi le pays. La conséquence est que cette première définition n'est pas suffisante pour distinguer le

vrai du faux dans de nombreuses situations. L'approche complémentaire choisie dans le présent travail est celle adoptée par plusieurs auteurs travaillant sur les FI (Nyhan & Reifler, 2010 ; Vlasceanu & Coman, 2022 ; Vraga & Bode, 2020), consistant à distinguer le vrai du faux sur la base des meilleures preuves disponibles de la part d'experts éminents. C'est pourquoi il sera admis ici de considérer comme vraies des propositions telles que : « *Le réchauffement climatique est réel et d'origine anthropique.* » (Catellani, 2021) ; « *Il n'y a pas de lien de cause à effet entre le vaccin RRO (Rougeole, rubéole et oreillons) et l'autisme.* » (Sénéchal et al., 2004) ; « *L'Irak, en 2003, ne disposait pas d'armes de destruction massive.* » (Deyries, 2022), bien que leur véracité soit parfois contestée (Cordonier et al., 2021 ; Lewandowsky et al., 2005, 2009 ; Ratzan, 2010) et qu'il n'est pas possible de strictement connaître leur correspondance avec l'état du monde. Cette conceptualisation est proposée en restant conscient que de nombreux débats entourent la définition du vrai et du faux (Davidson, 1996 ; Frápolli, 2009 ; E. Klein, 2006 ; Ludwig, 2016) et que d'autres approches auraient été possibles.

1.1.3 La notion de fausses informations

Les définitions précédentes nous amènent à définir les FI comme étant « des données contextualisées qui ne correspondent pas à l'état du monde ». Cette notion est à distinguer des termes connexes de désinformation, de mésinformation, de fausses nouvelles et de théories du complot (TC). Concernant les TC, il en existe de nombreux exemples, tels que les récits déclarant que le réchauffement climatique serait un mensonge pour faire avancer les politiques, que Bill Gates souhaiterait utiliser une campagne de vaccination massive contre le COVID-19 afin d'implanter des micropuces chez les personnes inoculées, ou que ce virus aurait été fabriqué intentionnellement dans un laboratoire dans le but d'éradiquer une partie de la population (voir Huchon & Schmidt, 2022 ; Reichstadt, 2019). Une définition des TC est qu'elles sont des « explications d'évènements spécifiques comme étant le résultat d'actions d'un petit groupe (i.e., les conspirateurs) qui agissent secrètement et cachent la vérité »

(Wagner-Egger & Gygax, 2018). Une première distinction entre les FI et les TC est que les TC ne sont pas strictement des informations mais des explications ou des récits. La nuance essentielle est qu'une information se réfère à un contenu spécifique tandis qu'une « théorie » englobe plusieurs contenus et y apporte une explication. Par exemple, déclarer qu'un médicament est dangereux alors qu'il ne l'est pas serait une FI, alors que l'idée selon laquelle un médicament dangereux est délibérément diffusé pour éliminer une partie de la population afin de servir les intérêts d'entités politiques relèverait d'une TC. La seconde distinction entre les TC et les FI est que les TC ne sont pas fausses par définition. L'histoire est en effet riche de complots réels (Wagner-Egger, 2021) et donc de TC qui ont pu se révéler vraies (voir cependant Dieguez & Delouée, 2021), tels que l'affaire du Watergate qui a montré que le président Nixon avait réellement comploté contre le parti démocrate en utilisant des procédés d'espionnage (Schudson, 2004). Malgré ces exceptions, il est important de souligner que les TC ont une probabilité élevée d'être erronées (Clarke, 2002 ; Wagner-Egger, 2021). Une raison majeure est qu'elles vont fréquemment à l'encontre du consensus d'experts, ce qui augmente par définition le risque qu'elles soient fausses. Par exemple, la théorie selon laquelle le réchauffement climatique serait un mensonge pour faire avancer les politiques s'oppose au consensus d'experts qui supporte la véracité de ce phénomène (Catellani, 2021). Il est par conséquent probable que la plupart des informations qui émergent des TC soient fausses et donc qu'un individu qui adhère à des TC adhère à des FI. C'est pour cette raison que dans le cadre de cette thèse, la littérature autour des TC sera parfois utilisée pour compléter le sujet des FI.

La notion de FI est aussi à nuancer par rapport à celle de fausses nouvelles. Les fausses nouvelles (aussi nommées « *fake news* ») sont définies comme étant « des informations clairement fabriquées qui imitent les médias d'information en ligne, dans la forme mais pas dans le contenu, et qui peuvent être politiques ou non » (Lazer et al., 2018 ; voir aussi Roussin, 2021). Élu mot de l'année par le dictionnaire britannique Collins en 2017 (Delevey, 2017), la

notion de *fake news* fait donc référence à un type particulier de FI se caractérisant par une forme sémantique, une origine médiatique, une large diffusion ainsi que par l'intentionnalité de leur création (Allard-Huver, 2017). Ainsi, un exemple de fausse nouvelle pourrait être un article de journal affirmant sans preuve que « *Hillary Clinton était ivre le soir de l'élection présidentielle* » (Pennycook et al., 2018).

Enfin, les FI ne sont pas de la mésinformation ni de la désinformation. La mésinformation et la désinformation désignent la production et/ou la diffusion de FI (Wardle, 2018). Elles constituent donc l'action dont les FI sont le résultat. Notons que ces deux termes se distinguent entre eux par leur intentionnalité. La mésinformation correspond à la diffusion accidentelle d'une FI (Treen et al., 2020 ; Wardle, 2018). A l'inverse, la désinformation se caractérise par l'intention de tromper (François, 2016 ; Huyghe, 2019 ; Klen, 2016). Concrètement, un exemple de désinformation serait l'acte de mensonge et une illustration de mésinformation serait de commettre une erreur dans la diffusion d'une information.

1.2 Importance de la problématique des fausses informations

1.2.1 Diffusion des fausses informations

Comme précisé précédemment (voir Avant-propos), plusieurs éléments de l'actualité invitent à penser que les FI sont une problématique d'importance actuelle. A notre connaissance, il n'existe pas de recherches évaluant la quantité de FI à notre époque et dans le temps qui étudieraient directement l'évolution de leur prévalence. Plusieurs évidences indirectes permettent toutefois de suggérer que leur nombre est important et en hausse.

1.2.1.1 Une diffusion importante

Un premier argument supportant l'importance de la diffusion des FI part du constat que la quantité totale d'informations disponibles est maintenant accrue, ce qui est le résultat de plusieurs étapes successives de l'histoire. Le développement du langage et de l'écriture puis de l'imprimerie, de la presse et de l'alphabétisation auraient participé à la diffusion, à l'accès et au

stockage des informations écrites. Ensuite, le développement de la télégraphie, suivi de la radio et de la télévision auraient permis aux informations orales de s'affranchir des frontières matérielles et de se multiplier (Dewar, 1998 ; Fang, 1997 ; Lamy & Petitjean, 2021). Avant les années 90, un effort restait toutefois nécessaire pour accéder, diffuser et stocker des informations. Elles étaient souvent payantes, leur diffusion limitée et leur stockage devait se faire de façon matérialisée. Une dernière révolution a alors été portée par Internet (le réseau physique, Mille, 2014) et le web (les documents numérisés accessibles avec un navigateur, Mille, 2014) qui se sont fortement développés à partir de la moitié des années 90 (Mille, 2014).

Internet et le web ont permis de réduire les limites à la diffusion des informations en leur assurant un accès majoritairement gratuit, une circulation internationale aisée et un stockage numérisé moins contraignant (Bronner, 2013). Le développement de ces outils coïncide ainsi avec une multiplication des informations disponibles. Au début des années 2000, des chercheurs auraient comptabilisé que l'information produite en cinq ans était quantitativement supérieure à celle produite depuis les débuts de l'imprimerie (Autret, 2002, cité dans Bronner, 2011). De l'an 2000 à 2015, le nombre d'articles scientifiques dans le monde serait passé de 800 000 à 1,8 millions (*Planetoscope - Statistiques : Publications d'articles scientifiques dans le monde*, s. d.). Dans le cadre des FI, étant donné que la quantité totale d'informations disponibles serait aujourd'hui accrue, la conséquence est qu'il est probable que la quantité absolue d'informations qui sont fausses se soit également développée.

En plus de cet argument statistique, un second élément qui supporte l'importance de la diffusion des FI est l'analyse des contenus sur le web (Lazer et al., 2018). Notamment, de nombreuses recherches observent que les FI se diffusent fortement sur les réseaux sociaux. Par exemple, en 2021 sur TikTok, réseau social le plus populaire et domaine le plus consulté (Tomé & Cardita, 2021), une recherche observait que près de 20 % des premiers résultats concernant des sujets d'actualité (e.g., guerre Russie-Ukraine, avortement, fusillades dans des écoles)

étaient des informations fausses (Brewster et al., 2022). De même sur YouTube, de nombreuses vidéos (Allgaier, 2019) et une partie notable des commentaires (Ha et al., 2022) participeraient à la diffusion de FI. Par exemple en 2020, plus d'un quart des vidéos les plus vues sur YouTube concernant le COVID-19 relatait de fausses informations (Li et al., 2020). De la même manière, en 2020 sur Twitter, une étude a analysé la véracité d'une multitude de tweets et a observé que de plus en plus semblaient contenir des FI (Vosoughi et al., 2018). Sur Facebook aussi, en 2020, une recherche conduite avec les mots-clefs *Zika* et *virus* relevait que, parmi les 200 premières publications proposées, 12 % étaient des FI (Sharma et al., 2017). Enfin, de façon générale, un sondage en 2024 a révélé que 74 % d'une population française affirmait être souvent exposé à des informations fausses sur les réseaux sociaux (IPSOS, 2024) (voir aussi Bode & Vraga, 2021).

1.2.1.2 Raisons à la diffusion

La littérature recense plusieurs explications à la diffusion importante des FI sur Internet. Une première résiderait dans les changements de la manière dont les informations diffusées sont maintenant vérifiées. Selon Bronner (2013, 2021), avant Internet, la plupart des informations diffusées étaient vérifiées car elles provenaient de médias professionnels. Internet aurait modifié la configuration de l'environnement informationnel en apportant une liberté nouvelle dans la diffusion d'informations. Un internaute quelconque peut maintenant créer un contenu, le publier sur le web et disposer d'un potentiel de rayonnement similaire à celui dont disposaient les médias. Cette liberté a pour conséquence que la plupart des informations diffusées ne font plus l'objet d'une vérification, augmentant ainsi la probabilité qu'elles soient inexactes. Supportant cette hypothèse, il est observé que ce sont les créateurs de contenu indépendants (comparés aux organisations certifiées par exemple) qui, sur YouTube, sont les plus à même de diffuser des informations fausses (Stocking et al., 2020). De même, une analyse des sources de FI sur Twitter observait que la plupart étaient des comptes d'internautes n'étant

pas publiquement connus (Vosoughi et al., 2018). De façon plus générale, une étude a examiné la provenance de 225 FI et les réseaux sociaux en constituaient la source pour 88 % d'entre elles (Brennen et al., 2020).

En plus de la moindre vérification des informations diffusées, une autre raison possible à la diffusion notable des FI sur Internet est d'ordre pécuniaire. Le nombre de visites sur un site Internet peut constituer une source de revenus et une manière d'augmenter le nombre de visites est probablement de proposer des informations attractives. Or, comparées aux informations vraies, les FI seraient généralement plus attractives. En effet, une recherche a observé que les FI se diffuseraient plus facilement et plus rapidement sur Twitter (Vosoughi et al., 2018). De même, en 2016, l'étude de Sharma et al. (2017) analysant les 200 premières publications Facebook données par une recherche comportant les mots *virus* et *Zika* a relevé que les publications contenant des FI (e.g., le virus Zika est un canular) étaient en moyenne plus partagées (740 partages en moyenne) par rapport à celles contenant des informations vraies (200) (Sharma et al., 2017). La conséquence de l'attractivité des FI est que leur diffusion constituerait une source de revenus supérieure aux vraies informations, ce qui motiverait leur diffusion. Ce phénomène s'est illustré en 2016 au cours des élections présidentielles américaines, où de nombreux créateurs de contenu auraient participé à la diffusion de FI afin de gagner de l'argent. En l'occurrence, des enquêtes ont identifié que la plupart des FI pro-Trump les plus partagées provenaient d'une petite ville en Macédoine du Nord et des entretiens menés par des journalistes avec leurs créateurs semblaient confirmer leur motivation pécuniaire (Hughes & Waismel-Manor, 2021 ; Subramanian, 2017).

1.2.2 Adhésion aux fausses informations

1.2.2.1 Une adhésion importante

En plus de leur diffusion marquée, un autre élément qui démontre l'importance de la problématique des FI est qu'elles reçoivent une adhésion élevée, comme le supportent les études

qui interrogent les croyances dans les TC. Comme spécifié précédemment, des exemples de TC sont les récits déclarant que le réchauffement climatique serait un mensonge pour faire avancer les politiques, que Bill Gates souhaite utiliser une campagne de vaccination massive contre le COVID-19 afin d'implanter des micropuces chez les personnes inoculées, ou que ce virus aurait été fabriqué intentionnellement dans un laboratoire dans le but d'éradiquer une partie de la population (voir Huchon & Schmidt, 2022 ; Reichstadt, 2019). Parmi ces trois théories, 10 % d'un échantillon international jugeait la première proposition « vraie » (Cordonier et al., 2021), plus de la moitié d'un échantillon américain identifiait la seconde comme tel (Sanders, 2020) et 15 % d'un échantillon français se déclarait en accord avec la troisième (Reichstadt & Fourquet, 2020). Dans la même idée, en 2019 (soit avant la pandémie de COVID-19), une enquête révélait que plus de 40 % des français était d'accord avec l'affirmation selon laquelle « *Le ministère de la Santé est de mèche avec l'industrie pharmaceutique pour cacher au grand public la réalité sur la nocivité des vaccins.* » (Reichstadt, 2019). De façon plus générale en 2019, plus d'un cinquième de la population française se déclarait en accord avec au moins cinq TC (Reichstadt, 2019) et plus d'un tiers de la population française affirmait « croire aux théories du complot » en 2023 (IFOP, 2023) (voir aussi CSA, 2022). En somme, l'analyse de l'adhésion aux TC montre que ces dernières sont fortement adoptées, ce qui permet d'inférer une adhésion significative à des FI au sein de la population.

1.2.2.2 Raisons à l'adhésion

La littérature recense plusieurs explications au niveau important d'adhésion aux FI. Une première réside dans l'utilisation des réseaux sociaux. En 2021, la moitié de la population française déclarait passer entre une heure et trois heures sur ces réseaux et la même proportion rapportait les utiliser comme sources d'information (CSA, 2022). De même, en 2020, presque trois quarts des américains affirmaient utiliser YouTube comme une source importante d'information (Stocking et al., 2020). Or, de nombreuses études observent qu'une unique

exposition à une information fausse est suffisante pour susciter de l'adhésion. Par exemple, l'exposition à une FI compromettant un politicien engendre chez les individus des attitudes plus négatives à son égard et envers son parti (Dobber et al., 2021), l'exposition à des FI faisant le lien entre le vaccin ROR et l'autisme provoque une baisse de la faveur envers cette vaccination (Featherstone & Zhang, 2020), et être informé (sans fondement) qu'il y aurait des vers dans les hamburgers de MacDonald réduit l'attrait pour ce restaurant (Misra, 1992). De ce fait, une explication possible à l'adhésion actuelle aux FI est simplement que les individus y sont fortement exposés du fait de leur utilisation importante des réseaux sociaux sur lesquels elles se diffusent (voir partie 1.2.1). Supportant cette explication, l'adhésion à des FI apparaît plus fréquente chez les individus qui utilisent les réseaux sociaux plusieurs fois par jour (Kraus et al., 2023) et qui leur font confiance en tant que source d'information (Freeman et al., 2020 ; Garrett, 2011 ; Melki et al., 2021). Par exemple, 29 % des utilisateurs de TikTok seraient d'accord avec la proposition « *L'assaut du Capitole en janvier 2021 a été mis en scène pour accuser les partisans de Donald Trump.* » contre 19 % de ceux qui ne l'utilisent pas (Kraus et al., 2023). De même, une autre expérience a interrogé le niveau d'adhésion à des FI relatives au COVID-19 ainsi que le degré d'utilisation des réseaux sociaux pour s'informer sur ce sujet et a relevé un lien positif entre les deux variables (Melki et al., 2021).

Une autre explication possible à l'adhésion actuelle aux FI peut s'expliquer à partir de la tendance qu'auraient les individus à privilégier les éléments du réel qui correspondent à leurs croyances (Larivée et al., 2018). Plusieurs recherches supportent l'existence d'un tel biais de confirmation (Chtourou, 2022 ; George, 1991 ; Kunda, 1990 ; Larivée et al., 2018, 2019 ; Sherman & Cohen, 2002). Par exemple, des individus confrontés à des arguments pour ou contre une TC ont tendance à juger plus persuasifs les arguments *pour* s'ils sont eux même en faveur de cette théorie et inversement s'ils sont en défaveur (McHoskey, 1995). Selon Bronner (2013, 2021), une explication à l'adhésion actuelle aux FI pourrait résider dans une interaction

entre ce biais de confirmation et la multiplication des contenus informationnels offerte par Internet. Avant cet outil, la plupart des informations diffusées étaient normalement vérifiées et leur contenu devait donc être relativement homogène. Par exemple, suivant les travaux d'experts, la plupart des sources d'information disponibles auraient convergé vers le consensus selon lequel le réchauffement climatique existe (Catellani, 2021). Par conséquent, la trajectoire d'un individu qui adhérerait à une fausse idée sur le climat aurait été de finir par rencontrer une convergence d'informations contradictoires. Une nouveauté avec Internet est que les FI sont plus facilement diffusées et accessibles. La conséquence est qu'un individu qui adhère à une fausse idée peut maintenant plus facilement s'engager avec succès dans une démarche confirmatoire. En ce sens, une recherche a par exemple observé que sur Facebook, un utilisateur qui *aime* les publications de pages complotistes a moins tendance à *aimer* les publications de pages scientifiques (Zollo et al., 2017). De la même manière, lors des élections américaines de 2016, les supporters de Donald Trump avaient plus tendance à consulter des faux articles pro-Trump et les pro-Clinton à consulter des faux articles favorables à Hillary Clinton (Guess & Nyhan, 2016). Le résultat de cette démarche confirmatoire facilitée est que les individus seraient moins souvent confrontés à des contradictions tout en rencontrant davantage d'informations confirmant ce à quoi ils croient.

Une troisième raison à l'adhésion aux FI pourrait s'opérer de façon plus passive, à travers le fonctionnement des algorithmes sur Internet, bien que l'impact de ces derniers soit discuté par certains auteurs (Guess et al., 2018 ; Robertson et al., 2018). Par le comportement en ligne d'un individu, les algorithmes sont censés déterminer ses préférences et ajuster en conséquence les éléments les plus pertinents à lui proposer en priorité (Farchy & Tallec, 2023 ; Robertson et al., 2018). Par exemple, une étude sur Facebook a observé les pages aimées par des individus et leurs amis ainsi que le contenu de leur fil d'actualités. Elle a alors relevé que si une page était aimée par un individu ou par un de ses amis, alors le contenu de cette page

avait plus de probabilités d'apparaître sur son fil d'actualités (Boyadjian, 2020). Dans le cadre des FI, la conséquence de ce fonctionnement est qu'un individu démontrant sur Internet un intérêt pour une information fausse, par ses recherches, ses amis ou les pages auxquelles il s'abonne, devrait ensuite être plus souvent exposé à des éléments qui la confirment. À terme, ce fonctionnement pourrait renforcer l'adhésion à la FI, en l'exposant davantage à cette information et à des éléments qui la confirment (Houston, 2018 ; Novin & Meyers, 2017). En somme, le fonctionnement des algorithmes en plus de la multiplication des contenus informationnels et de la diffusion aisée des FI constituent trois explications, complémentaires et non-exhaustives, à l'adhésion importante qu'elles paraissent récolter.

1.3 Ubiquité des conséquences des fausses informations

Favorisées par le fonctionnement d'Internet, les FI semblent donc disposer d'une diffusion et d'une adhésion significative. Ce phénomène suscite des répercussions notables dans plusieurs domaines de la société actuelle (Goertzel, 2010 ; Huchon & Schmidt, 2022) tels que celui de la santé. En oncologie par exemple, des informations peuvent affirmer que des thérapies alternatives (e.g., jus de fruits) seraient plus appropriées que des traitements allopathiques. De nombreuses FI existent également autour du sujet de la vaccination, comme l'ont illustré les contenus autour du vaccin du COVID-19 (Bottemanne, 2022) ou encore le cas connu d'un article publié par Andrew Wakefield. En 1998, cet ancien médecin a publié un travail qui affirmait l'existence d'un lien entre la vaccination Rougeole – Oreillons – Rubéole (ROR) et le trouble du spectre autistique (Ratzen, 2010). Quelques années plus tard, les résultats de l'article se sont avérés faux voire falsifiés (Poland & Spier, 2010 ; Sénéchal et al., 2004). Le problème majeur avec ces FI dans le domaine sanitaire est qu'elles sont dangereuses car elles peuvent impacter négativement la santé des personnes qui y adhèrent. Une étude a par exemple interrogé des individus souffrant de cancer et a relevé que ceux qui adhéraient à des FI autour des traitements oncologiques étaient plus enclins à refuser des traitements conventionnels en

faveur de thérapies alternatives (Citrin et al., 2012). Or, retarder ou refuser la prise en charge par des traitements allopathiques en faveur de thérapies alternatives est associé à un plus haut risque de mortalité (E. Y. Chang et al., 2006). Dans un autre registre, il peut être noté qu'une hausse de cas de rougeole et une réduction de la couverture vaccinale ROR coïncident avec l'année de publication de l'article d'Andrew Wakefield (Jansen et al., 2003 ; Sauvayre, 2021). De façon plus expérimentale, une expérience a relevé que l'exposition à des FI faisant le lien entre le vaccin ROR et l'autisme provoquait une baisse de la faveur envers cette vaccination (Featherstone & Zhang, 2020). Enfin, de nombreuses études ont étudié l'adhésion aux FI sur les vaccins et ont observé qu'elle prédisait une moindre tendance à accepter les recommandations gouvernementales, la démarche diagnostique et la vaccination (Freeman et al., 2020 ; Jolley & Douglas, 2014 ; Loomba et al., 2021 ; Roozenbeek, Schneider, et al., 2020).

Des FI peuvent aussi impacter le sujet de l'environnement et du climat, en niant par exemple l'existence du changement climatique ou son explication anthropogène (i.e., le changement climatique serait causé par l'Homme) (Treen et al., 2020). Plusieurs recherches ont montré qu'une simple exposition à ce type d'information pouvait réduire la propension des individus à croire en un consensus scientifique sur le changement climatique ou à penser que ce dernier était causé par l'Homme (Drummond et al., 2020 ; Porter et al., 2019 ; Ranney & Clark, 2016 ; van der Linden, 2015). L'adhésion à ces informations fausses a des répercussions notables, puisqu'il est observé qu'elle réduit la probabilité d'adopter des comportements pro-environnementaux (van der Linden, 2015). Notons qu'en 2017, plus d'un quart des Américains estimait que le réchauffement climatique n'était pas réel et presque la moitié jugeait que sa cause n'était pas humaine ou qu'il n'y avait pas de consensus scientifique à son sujet (Ballew et al., 2019).

Dans le domaine de la justice, l'exposition à de fausses informations peut également s'avérer problématique. Par exemple, des jurés peuvent être exposés à des aveux forcés (Kassin

& Sukel, 1997), à des faux témoignages ou à des preuves falsifiées (Black, 2018 ; Fein et al., 1997 ; Wolf & Montgomery, 1977), ce qui peut avoir des conséquences majeures. Plusieurs recherches observent en effet que ces FI peuvent modifier le jugement des jurés, augmentant par exemple leur tendance à déclarer que l'accusé est coupable si les fausses évidences l'incriminent (Black, 2018 ; Fein et al., 1997 ; Kassin & Sukel, 1997 ; Wolf & Montgomery, 1977).

Le domaine politique est également exposé aux FI (Van Bavel et al., 2021). Sur Twitter, une recherche observait d'ailleurs que c'est ce type de FI qui se diffuserait de la façon la plus rapide et la plus large, comparativement à d'autres sujets (e.g., terrorisme, catastrophes naturelles, science) (Vosoughi et al., 2018). Des illustrations connues de FI politiques aux Etats-Unis sont des informations accusant Barack Obama de ne pas avoir la nationalité américaine (Tuchman, 2011) ou d'autres proposant que Hillary Clinton aurait vendu des armes à l'Etat Islamique (Hughes & Waismel-Manor, 2021). Les FI politiques seraient particulièrement prégnantes lors de campagnes présidentielles (Winneg et al., 2005). Un mois avant l'élection présidentielle américaine de 2016, il est estimé que la plupart des citoyens américains auraient été exposés à au moins une fausse nouvelle (Allcott & Gentzkow; 2017) (voir aussi Zimmermann & Kohring, 2020). Les FI politiques sont une problématique cruciale car, en plus d'être associées à une hausse de la violence (Jolley & Paterson, 2020), elles peuvent influencer les attitudes politiques de ceux qui les reçoivent (Allcott & Gentzkow, 2017 ; voir cependant Karpf, 2019). Par exemple, l'exposition à une FI compromettant un politicien semble engendrer chez les individus des attitudes plus négatives à son égard et envers son parti (Dobber et al., 2021). De plus, cette influence des FI politiques pourrait aller jusqu'à impacter les comportements ou intentions de vote (Weeks & Garrett, 2014 ; Zimmermann & Kohring, 2020). En effet, une étude conduite après les élections présidentielles américaines de 2008 (Obama vs McCain) a relevé que le nombre de rumeurs sur Obama auquel avaient été exposés

des individus démocrates prédisait négativement leur choix de vote en sa faveur (Weeks & Garrett, 2014). Également, le degré d'adhésion à des théories complotistes islamophobes permettait de prédire les intentions de vote de citoyens anglais lors du Brexit (Swami et al., 2018). Plusieurs auteurs soulignent par ailleurs que l'influence des FI sur les attitudes politiques est d'autant plus grave dans les sociétés démocratiques qui aspirent à ce que le résultat d'un vote soit le résultat d'un processus éclairé de la part des citoyens, et pas d'un raisonnement tronqué par des informations fausses (Van Bavel & Pereira, 2018 ; F. Webster, 1999).

Le domaine militaire est aussi soumis à de nombreuses FI. Durant la Première Guerre mondiale (voir Bloch, 1921), de fausses informations diffusées en Angleterre auraient affirmé que les Allemands usaient des cadavres de leurs soldats pour nourrir leurs animaux (Posetti & Matthews, 2018). A la même époque, le gouvernement britannique aurait fait passer des navires armés anglais comme étant des bateaux de commerce inoffensifs afin d'attirer à la surface les sous-marins allemands et de pouvoir ainsi les abattre (Coder, 2000). Plus tard lors de la Seconde Guerre mondiale, l'Opération Fortitude a consisté en une opération de désinformation qui a employé des agents doubles et construit des tanks, des chalands ainsi que des antennes en plastiques dans le sud-ouest de l'Angleterre tout en y simulant un trafic de communications important, dans l'ambition de faire croire en la préparation d'un débarquement dans le Pas-de-Calais (Anderson, 2010 ; De Witte Parra, 2019 ; Kornegay, 2013 ; Tavares, 2001). Le 21e siècle est aussi riche d'exemples de FI militaires. En 2015 par exemple, des soldats ukrainiens auraient reçu des messages d'émetteurs inconnus leur annonçant que leur chef les avait trahis (Lebret, 2015). De même en 2020, l'OTAN a détecté plusieurs attaques informationnelles contre la présence de ses troupes en Lettonie, en Lituanie et en Pologne. Un faux courrier annonçait par exemple au ministre de la Défense lituanien que l'OTAN avait l'intention de retirer ses troupes du pays (NATO, 2020). Outre ces actions de désinformations, des faits de mésinformation sont aussi notables dans le domaine militaire. En 1988 par exemple,

dans la tragédie du vol Iran Air 655, le radar aérien d'un croiseur américain aurait par erreur indiqué qu'un aéronef qui l'approchait était un avion de combat ennemi (Rochlin, 1991). Toutes les FI dans le domaine militaire constituent un enjeu particulièrement important pour les armées qui ont besoin d'une qualité de jugement et informationnelle optimales (Bellais, 1998 ; Biermann et al., 2004 ; Dzindolet et al., 2001), au risque de mener à des conséquences majeures voire tragiques. Par exemple, dans la tragédie du vol Iran Air 655, la défaillance du radar aurait participé à la décision du commandant du navire d'ouvrir le feu, avant de réaliser que sa cible était finalement un avion de ligne transportant 290 civils (Rochlin, 1991).

Enfin, les FI auraient aussi des répercussions conséquentes sur le plan économique. Une première concerne les fausses publicités positives ou négatives qui pourraient influencer le comportement des consommateurs. Par exemple, la présentation d'une fausse publicité positive sur un supermarché améliorerait l'attitude face à celui-ci (Misra, 1992). Également, le visionnage d'une publicité mensongère pour un bain de bouche indiquant qu'il serait capable de soigner les rhumes augmenterait la croyance que ce produit aurait cette qualité thérapeutique (Armstrong et al., 1979). En plus de la question des publicités mensongères, une autre préoccupation économique avec les FI est qu'elles peuvent avoir des conséquences dans le domaine de la finance. Dans son livre, Huyghe (2016) décrivait la fausse nouvelle selon laquelle le PDG de l'entreprise Apple aurait eu un malaise cardiaque. En moins d'une heure, cette information aurait fait chuter le cours de son entreprise de 10 %. De même, une FI affirmant la faillite de United Airlines aurait entraîné une chute de son cours boursier de plus de deux tiers de sa valeur (« United Airlines dément les rumeurs de faillite », 2008). Pour finir, une autre problématique économique avec les FI est qu'elles susciteraient des dépenses. Par exemple, étant donné que les FI sur le vaccin réduisent le taux d'inoculation (Freeman et al., 2020 ; Jolley & Douglas, 2014 ; Loomba et al., 2021) et que ce dernier réduit la prévalence de maladies, la moindre vaccination induite par les FI augmenterait les dépenses de santé (Mayers, 2019). De

la même manière, en France en 2020, il est possible que les croyances non-fondées envers la dangerosité de la 5G (Jolley & Paterson, 2020) aient contribué au vandalisme de plusieurs centaines d'antennes relais (Berthelot & Eliakim, 2021). Enfin, de dernières dépenses liées aux FI concernent les actions politiques engagées pour lutter contre leur influence (Cavazos, 2019).

1.4 Approche corrective face aux fausses informations

Parmi les actions politiques engagées, une partie est destinée au développement de la correction des FI (Cavazos, 2019). Dans le cadre de ce travail, nous utiliserons la notion d'approche corrective pour se référer aux actions dont l'objectif est de rétracter ou de réfuter des FI. Cette approche corrective peut être menée à plusieurs échelles. Par exemple, au niveau des institutions politiques, l'OTAN cherche à réfuter les FI d'autres pays (NATO, 2020), telles que les « fausses affirmations » russes à son égard en reprenant chacune d'entre elles et en rappelant « les faits » (OTAN, 2020, 2022). De même, l'Union Européenne a mis en place un groupe de travail spécifiquement dédié à lutter contre la désinformation du Kremlin (<https://euvsdisinfo.eu/>). Elle a aussi fondé en 2020 l'European Digital Media Observatory (<https://edmo.eu>) dont le projet est de regrouper des experts et des organisations spécialisés dans la désinformation afin d'optimiser la recherche et les projets sur ce sujet. La même année, l'ONU lançait l'initiative « Verified », dont l'objectif est de « couper court aux *fake news* » (<https://shareverified.com>).

Cette approche corrective peut aussi se concrétiser à d'autres échelles telles que l'intervention d'une personne qui en corrige une autre. Sur Twitter par exemple, des recherches illustrent qu'un internaute peut porter à la connaissance d'un pair que ce qu'il partage est infondé (e.g., Mosleh et al., 2021 ; Vraga & Bode, 2017a). En mars 2020, plus d'un tiers d'un échantillon d'une population d'internautes exprimant avoir été confronté à des FI autour du COVID-19 déclarait en avoir corrigé (Bode & Vraga, 2021). Également, la moitié de ceux qui

admettaient avoir partagé une FI sur ce sujet rapportaient avoir reçu une correction (Bode & Vraga, 2021).

Outre les internautes et les institutions politiques, des acteurs majeurs dans la correction des FI sont les organismes de vérification des faits (*fact-checking*). La plupart d'entre eux sont des médias d'information (Graves & Cherubini, 2016). Par exemple, l'Agence France Presse (AFP) a créé le service AFP Factuel (<https://factuel.afp.com/>). De même, le site Politifact (<https://www.politifact.com>) est un site entièrement dédié à la vérification d'informations et à leur correction le cas échéant. L'objectif principal des organismes de vérification des faits serait d'examiner la véracité de déclarations publiques afin d'améliorer la vie en société (Amazeen et al., 2018 ; Graves & Cherubini, 2016) et d'aider les citoyens à faire des choix éclairés (Amazeen, 2015 ; Elizabeth, 2014). Au total, 90 % du travail régulier des organisations de *fact-checking* en Europe serait dédié aux déclarations de politiciens (Graves & Cherubini, 2016), espérant ainsi réduire les biais dans les comportements de vote (Luengo & García-Marín, 2020). Soulignons que ces organismes ont connu une croissance importante au cours des années 2010. Leur chiffre aurait triplé de 2014 à 2018 (Stencel & Griffin, 2018) et doublé de 2016 à 2022 (Stencel et al., 2022). En 2022, ces organisations seraient au moins 350 à travers le monde (Ryan, 2022). Ainsi, soulignant l'importance de l'approche corrective, lors des dernières semaines des élections américaines de 2016, 25 % des américains déclaraient avoir été exposés à un contenu de vérification des faits (Guess & Nyhan, 2016).

En résumé, de nombreux acteurs, dont des organismes de vérification des faits, se mobilisent pour réfuter les affirmations erronées. Leur action apparaît nécessaire pour contrer l'influence des FI qui, bien qu'elles aient toujours existé, semblent aujourd'hui avoir une diffusion importante, susciter une adhésion significative et produire de nombreux impacts au sein de la société.

Chapitre 2. Les fausses informations persistent malgré leur correction

L'efficacité de l'approche corrective repose sur un modèle de déficit de la connaissance (Simis et al., 2016) dont le postulat est que les individus agiraient à tort parce qu'ils sont mal informés (Kuklinski et al., 2000). Le corollaire de ce postulat est qu'une façon de rectifier un comportement inapproprié serait simplement d'apporter une connaissance, laquelle serait traitée de façon rationnelle et objective. Cependant, toute la littérature autour de l'effet d'influence continue (Johnson & Seifert, 1994) démontre que l'influence d'une FI peut persister même quand elle est signalée comme étant fausse. L'objet de ce chapitre sera de définir cet effet avant d'en proposer des explications. Les autres résultats possibles à la suite de la correction d'une FI seront ensuite mentionnés.

2.1 Conceptualisation classique de l'effet d'influence continue

« *Simply stated, there is no doubt that Saddam Hussein now has weapon of mass destruction. There is no doubt he is amassing them to use against our friends, against our allies, and against us.* » Exprimée par Dick Cheney, vice-président des Etats-Unis en 2002, cette déclaration illustre le contenu de nombreuses informations diffusées à l'époque, qui avançaient que l'Irak détenait des armes de destruction massive. Les faits ont ensuite réfuté la véracité de ce discours puisqu'aucun armement de la sorte n'a été trouvé suite à l'invasion du pays. Malgré cela, des années plus tard, une partie significative de la population américaine continuait de croire en la présence de telles armes alors même qu'ils connaissaient la correction (Lewandowsky et al., 2005, 2009).

2.1.1 Définition

La notion d'Effet d'Influence Continue (« *continued influence effect* ») (EIC, Johnson & Seifert, 1994) est utilisée pour désigner le phénomène au sein duquel l'influence d'une FI persiste sur les cognitions malgré sa correction. Cet effet s'observe dans une situation composée de quatre temps majeurs : avant la réception de la FI, à la réception de la FI, à la réception de la correction puis après. Avant la réception de la FI, un individu peut avoir une évaluation

prédéfinie envers un objet, ou bien une absence d'avis. Ensuite, il recevrait une information et ses cognitions devraient normalement s'accorder avec cette information. Dans un troisième temps, une correction indiquant à l'individu que l'information précédente était fausse serait reçue. A la fin, un EIC serait observé si les cognitions continuent de s'accorder davantage avec la FI qu'elles ne l'étaient lors de la première étape. Par exemple, un individu pourrait n'avoir aucun jugement prédéfini sur la présence potentielle d'armes de destruction massive Irak. S'il lui est ensuite indiqué que l'Irak détenait des armes de destruction massive, alors il devrait croire que l'Irak détenait des armes de destruction massive (FI). Après quoi, il pourrait apprendre que les informations qu'il avait reçues sur l'Irak étaient fausses (correction). Dans la dernière étape, un EIC serait relevé si l'individu trouve plus plausible l'idée que l'Irak cache des armes de destruction massive par rapport à avant la réception de la FI.

Il est possible de distinguer deux formes possibles d'EIC. Dans une première, la correction n'a pas d'efficacité, c'est-à-dire que le niveau d'adhésion à la FI après la correction reste similaire à celui qu'il était avant la correction. Par exemple, après la correction, l'adhésion à la croyance selon laquelle l'Irak détenait des armes de destruction massive serait toujours aussi forte après la correction par rapport à avant. Dans une deuxième forme possible d'EIC, la correction a une influence mais elle reste insuffisante pour retourner au niveau de base. Par exemple, à la suite de la correction, l'adhésion à la croyance selon laquelle l'Irak cachait des armes de destruction massive pourrait passer de forte à modérée, mais resterait toujours plus élevée par rapport à avant la réception de la FI. Ces deux cas de figure sont illustrés au sein de la Figure 1.

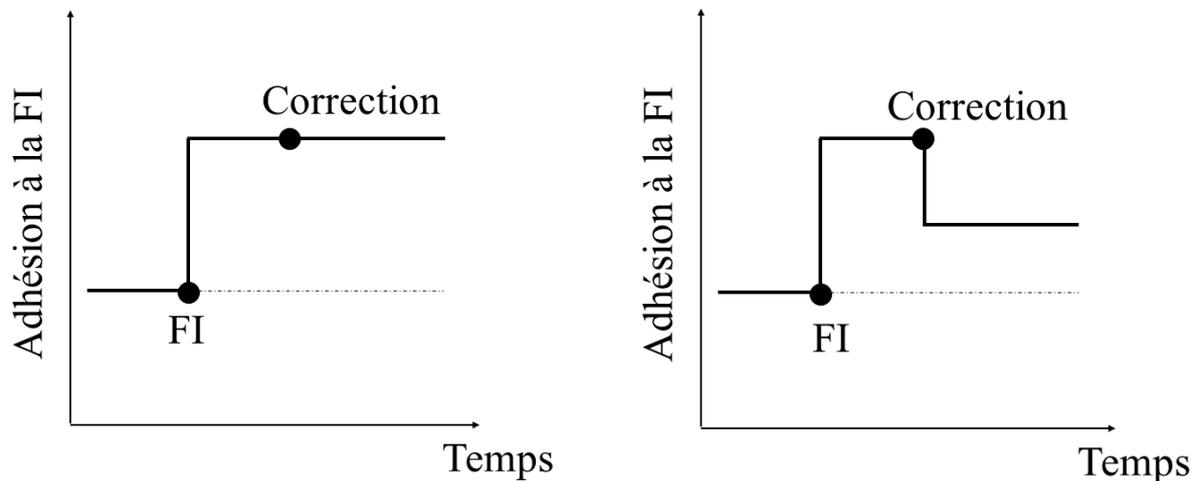


Figure 1. Correction inefficace (à gauche) et correction insuffisante (à droite)

2.1.2 Historique et paradigme classique

Un cadre théorique et une définition de l'EIC ont la première fois été proposés en 1994 par Johnson et Seifert. Leur article proposait cinq expériences observant un EIC sur des étudiants en psychologie aux Etats-Unis. Les participants étaient reçus en groupe et obtenaient en échange des crédits pour un cours. Ils recevaient un livret individuel contenant plusieurs pages, chacune comportant un message qu'ils avaient pour consigne de lire à leur rythme, comme s'ils lisaient un article de journal. Ils étaient informés qu'ils auraient à répondre à des questions sur le texte après sa lecture. Dans une expérience fréquemment citée (Johnson & Seifert, 1994, étude 1A), le texte, repris de Wilkes et Leatherbarrow (1988), décrivait l'intervention de pompiers sur un incendie se propageant dans un entrepôt. Les participants étaient informés que l'incendie aurait été déclenché par un court-circuit dans un cagibi dédié à des installations électriques. Pour un groupe de participants (groupe FI), l'un des premiers messages, dont la source était un policier, spécifiait que des matériaux inflammables avaient négligemment été stockés dans ce cagibi. Cette information était ensuite corrigée par le policier dans un autre message révélant que le cagibi était finalement vide. Un autre groupe de participant (groupe contrôle) recevait directement la vraie information (i.e., le cagibi était vide) et n'avait pas connaissance de la FI. Les résultats montraient alors que, malgré la correction,

les participants qui avaient reçu la FI continuaient de se référer davantage à des notions de matériaux inflammables et de négligence pour expliquer le texte par rapport au groupe contrôle (Johnson & Seifert, 1994, étude 1A). La taille d'effet ($d = 1,521$) était large (Lee, 2016 ; Maher et al., 2013). Dans le cadre de cette thèse, la notion de « texte d'EIC » sera utilisée pour faire référence à ce type de texte employant une FI subséquemment corrigée.

Dans l'expérience de Johnson et Seifert (1994), la variable dépendante qui était utilisée pour étudier cet effet était la réponse à des questions inférentielles. Les questions inférentielles évaluent les inférences des participants, c'est-à-dire des constructions mentales produites à partir d'informations explicites (Martins & Le Bouëdec, 1998). Par exemple, un individu qui lit qu'une personne a une chambre en désordre réaliserait une inférence s'il déduit qu'elle est maladroite ou qu'elle aurait tendance à oublier un objet dans un bus (Rapp & Kendeou, 2007). Concrètement, les questions inférentielles utilisées par Johnson et Seifert (1994) étaient des questions ouvertes telles que « *Sur quel aspect de l'incendie la police pourrait souhaiter continuer d'enquêter ?* ». Les auteurs examinaient alors si le nombre d'inférences produites en lien avec la FI (e.g., « *La police pourrait continuer d'enquêter sur de la négligence.* ») était différent en fonction des groupes.

Le paradigme proposé par Johnson et Seifert (1994) peut être considéré comme étant le paradigme classique de l'EIC. Pour cause, en plus de son caractère précurseur, leur travail est très fréquemment cité pour évoquer la persistance des FI malgré leur correction. De plus, plusieurs de ses caractéristiques principales sont reprises dans de nombreux travaux appartenant à cette littérature. En effet, beaucoup emploient un texte écrit constitué d'une série de messages et comparent les réponses à des questions inférentielles ouvertes entre un groupe contrôle sans FI et un groupe ayant reçu une FI corrigée (e.g., Brydges et al., 2018 ; Ecker, Lewandowsky, et al., 2014 ; Ecker et al., 2017 ; Ecker & Antonio, 2021 ; O'Rear & Radvansky, 2020).

Enfin, bien que l'EIC a clairement été défini en 1994 (Johnson & Seifert, 1994), il doit être relevé que de précédents travaux avaient déjà indirectement abordé la question de la persistance des FI malgré leur correction. A notre connaissance, la première expérience observant un effet analogue a été conduite en 1967 (Walster et al., 1967). Des sujets participaient à un exercice d'entretien censé évaluer leurs compétences sociales. Un résultat indépendant de leur vraie performance leur était d'abord donné, avant de révéler son caractère factice. Pour autant, les analyses montraient que les participants qui avaient reçu une note médiocre continuaient de s'attribuer une compétence sociale moindre par rapport à ceux qui avaient reçu une bonne évaluation (Walster et al., 1967). Plus tard, en 1975 (L. Ross et al., 1975), une expérience similaire a exposé des participants à un feedback aléatoire (réussite ou échec) sur une tâche qu'ils venaient de réaliser. Ils étaient ensuite informés du caractère randomisé du retour qui leur avait été fait. Malgré cela, les participants qui avaient entendu qu'ils avaient réussi avaient plus tendance à croire qu'ils avaient produit une bonne performance par rapport à ceux ayant appris qu'ils avaient échoué (voir aussi Misra, 1992).

2.2 Explications théoriques

La littérature offre de nombreuses explications à l'EIC (e.g., Connor Desai et al., 2020 ; Ecker et al., 2011 ; Ithisuphalap et al., 2020 ; Susmann & Wegener, 2021). A notre connaissance, ces différentes explications ont toujours été rapportées de façon éparse et n'ont jamais fait l'objet d'une synthèse complète. Cette partie a donc été rédigée avec l'objectif de réaliser une synthèse inédite des différentes explications les plus souvent rencontrées. Nous proposons de les diviser en trois approches : cognitive, motivationnelle et rationnelle.

2.2.1 Approche cognitive

Un premier cadre théorique autour de l'EIC repose sur des explications de nature cognitive. De manière générale, plusieurs expériences démontrent une implication des ressources cognitives dans cet effet. Par exemple, une recherche a exposé des participants à une

charge cognitive consistant à retenir une série de nombres pendant qu'ils lisaient un texte d'EIC. Un groupe de participants recevait la charge cognitive au moment de lire la FI et un autre lorsqu'il lisait la correction. Démontrant une implication des ressources cognitives, il était observé que la manipulation de ces dernières faisait varier l'EIC. En l'occurrence, l'EIC augmentait lorsque la charge était fournie au moment de la correction et réduisait quand elle l'était lors de la réception de la FI (Ecker et al., 2011 ; voir aussi Gilbert et al., 1990 ; Sanderson et al., 2022).

2.2.1.1 Modèle de la révision du modèle mental

Un modèle théorique qui permet de conceptualiser comment la sphère cognitive peut impacter la persistance d'une FI est le *Knowledge Revision Components* (KReC, Kendeou et al., 2014). Son postulat est que la lecture d'un écrit consisterait en l'intégration cyclique, en mémoire (O'Brien et al., 1998), d'une succession de propositions participant à la représentation de la situation décrite (Rapp & van der Broek, 2005). En somme, l'individu construirait un modèle mental de la situation qu'il perçoit, qui s'adapterait de façon dynamique et cohérente avec les informations qu'il reçoit (Johnson-Laird et al., 2012).

La structure d'un modèle mental serait toutefois remise en cause lorsque la correction d'une information est reçue. Pour cause, une correction implique de modifier une information au sein du modèle mental et donc de le mettre à jour. Le KReC (Kendeou et al., 2014) identifie cinq principes qui définiraient la révision d'un modèle mental à la suite de la correction d'une information : l'encodage, l'activation passive, la coactivation, l'intégration et la compétition d'activation. Selon le premier principe, à partir du moment où une information a été encodée, elle serait stockée en mémoire et ne pourrait plus être totalement effacée. Le second principe ajoute qu'une information encodée aurait toujours le potentiel d'être activée en mémoire si un stimulus de l'environnement qui lui est associé est perçu. La notion d'activation passive fait référence à l'idée que cette activation se ferait de façon automatique et passive. En l'occurrence,

lors de la réception d'une correction, la FI devrait être activée étant donné que les deux informations sont liées par une relation de conflictualité (e.g., le cagibi contenait des matériaux inflammables – le cagibi était vide). La conséquence de cette activation passive de la FI au moment de la réception de la correction serait une situation de coactivation (troisième principe) entre les deux informations, c'est-à-dire que la FI et la correction seraient simultanément activées. A partir de cette coactivation, l'individu pourrait mettre à jour son modèle mental en intégrant que la FI est fausse (quatrième principe). Selon le KReC (Kendeou et al., 2014), la coactivation serait toutefois une étape préalable nécessaire pour que le processus d'intégration puisse se réaliser. Autrement dit, il serait nécessaire que le conflit entre la correction et la FI soit détecté pour que le modèle mental puisse ensuite être mis à jour. Enfin, le dernier principe (compétition d'activation) stipule que, même si la correction a été intégrée au sein du modèle, la FI pourrait toujours être activée lorsque le modèle mental est sollicité. Par exemple, dans l'étude de Johnson et Seifert (1994), une question telle que « *Quelle est la cause de l'incendie ?* » devrait toujours activer la notion de matériaux inflammables, même si les individus ont intégré qu'il n'y avait finalement pas de tels matériaux dans le cagibi. La conséquence de cette activation simultanée entre la correction et la FI au moment de répondre à une sollicitation du modèle est que la réponse à cette sollicitation dépendra du résultat de la compétition d'activation entre la correction et la FI. Ce dernier point sera développé dans la partie 2.2.1.3.

Selon les auteurs, la validité du KReC (Kendeou et al., 2014) est notamment supportée par les études qui montrent qu'une même information est lue plus lentement lorsqu'elle contredit une information précédemment présentée par rapport à si elle n'en contredit pas (Kendeou et al., 2013 ; O'Brien et al., 1998 ; Rapp & Kendeou, 2007 ; Van Den Broek & Kendeou, 2008). Par exemple, dans une expérience plusieurs fois répliquée (O'Brien et al., 1998), des participants ont lu un texte dans lequel un individu commandait un cheeseburger.

Cette dernière information était lue plus lentement si le texte avait précédemment indiqué qu'il était végétarien. De même dans un autre paradigme, un temps de lecture plus long était observé pour un énoncé scientifique s'il allait à l'encontre des croyances initiales des participants (Van Den Broek & Kendeou, 2008). Selon les auteurs, ces résultats supportent la validité du KReC car l'activation passive de la FI à la lecture de la correction, la détection du conflit entre la FI et la correction ainsi que l'intégration de la correction au sein du modèle devraient susciter des interférences dans la lecture et donc un temps de lecture plus lent.

A partir du KReC, il est possible de distinguer deux moments propices à un EIC. Le premier est au moment de la réception de la correction. A ce moment, un EIC pourrait se produire si le modèle mental n'est pas mis à jour malgré la réception de la correction. Le second est lorsque le modèle mental est sollicité, avec une question inférentielle par exemple. A cet instant, un EIC pourrait s'observer si la correction est insuffisamment activée. Ces deux moments sont développés ci-après.

2.2.1.2 Défaut de mise à jour du modèle mental

Un premier moment propice à la survenue d'un EIC est le moment de la réception de la correction. Ici, un EIC pourrait se produire en raison d'un défaut de mise à jour du modèle mental (Jia et al., 2020 ; Kendeou et al., 2014 ; Rapp & Kendeou, 2007). Ce défaut de mise à jour du modèle mental pourrait survenir par deux processus. Le premier est un défaut d'activation de la FI et le second est un défaut d'intégration de la correction à la place de la FI.

2.2.1.2.1 *Défaut d'activation de la fausse information*

Une première situation qui pourrait mener à un défaut de mise à jour du modèle mental est un défaut d'activation de la FI lors de la réception de la correction. Un défaut d'activation de la FI est possible si son conflit avec la correction n'est pas détecté. Par exemple, dans l'expérience de Johnson et Seifert (1994), un participant pourrait lire que le cagibi était vide, sans pour autant prendre conscience que cela implique une absence de matériaux inflammables

dans la pièce. En l'absence de détection du conflit entre la FI et la correction, alors la FI ne serait pas activée. Par conséquent, la correction ne pourrait pas être intégrée au sein du modèle à la place de la FI, résultant en une persistance de son influence (Van Den Broek & Kendeou, 2008).

Le rôle d'un défaut d'activation de la FI au moment de la réception de la correction sur l'EIC est supporté par plusieurs travaux. Par exemple, dans une expérience (Ithisuphalap et al., 2020), des participants ont été exposés à plusieurs textes d'EIC. Durant la lecture, un groupe de participants était invité à évaluer leur croyance en une information qui s'apprêtait à être corrigée. Les résultats ont montré que l'évaluation de la croyance en la FI avant sa correction réduisait l'EIC par rapport à un groupe n'ayant pas reçu cette consigne. L'interprétation des auteurs était que ce moment de réflexion sur la FI l'avait rendue plus disponible en mémoire, ce qui avait donc facilité son activation au moment de l'encodage de la correction. Par ailleurs, une autre recherche (Rich & Zaragoza, 2015) a confronté des participants à un texte d'EIC relatif à un cambriolage, pour lequel un suspect était proposé (FI) avant qu'il ne soit innocenté (correction). Dans une condition, l'implication du suspect était explicitement mentionnée dans le texte alors qu'elle ne l'était qu'implicitement dans une autre. Les résultats ont montré que l'EIC était moins important lorsque la FI avait été explicitement mentionnée. En cohérence avec l'idée d'un défaut de coactivation, une explication possible est qu'une information explicite serait plus facilement accessible en mémoire et donc plus aisément réactivée au moment de la réception de la correction (Rich & Zaragoza, 2015). Enfin, une autre étude a observé une plus grande réduction de l'EIC lorsque la FI était mentionnée dans la correction par rapport à si la vraie information était mentionnée sans reprendre explicitement la FI (Ecker et al., 2017). Ce résultat est également cohérent avec l'hypothèse de l'effet d'un défaut d'activation de la FI au moment de la correction sur l'EIC, puisque la mention de la FI dans la correction devrait faciliter la coactivation entre la FI et la correction.

2.2.1.2.2 *Défaut d'intégration de la correction*

Une seconde situation qui pourrait mener à un défaut de mise à jour du modèle mental est un défaut d'intégration de la correction. Pour rappel, le principe d'intégration stipule que pour mettre à jour un modèle mental à la suite de la correction d'une information, le sujet doit remplacer la FI par la correction. Pour être réalisée avec succès, l'intégration nécessiterait toutefois des ressources cognitives. Des ressources cognitives sont nécessaires afin d'activer le modèle mental, prendre conscience du conflit entre la correction et la FI, intégrer le caractère faux de la FI, tout en traitant les nouvelles informations provenant de l'environnement extérieur. Par conséquent, si les ressources cognitives sont insuffisantes au moment de la réception de la correction, alors cette insuffisance pourrait mener à un défaut dans la mise à jour du modèle mental et donc à une persistance de la FI. Par exemple, dans l'expérience de Johnson et Seifert (1994), même si un participant a bien détecté le conflit entre la présence de matériaux inflammables dans le cagibi et l'information selon laquelle le cagibi était vide, le modèle mental pourrait ne pas être mis à jour si les ressources cognitives disponibles ne lui permettent pas de réaliser cette opération.

Conformément à l'hypothèse du défaut d'intégration de la correction sur l'EIC, il est observé qu'une FI persiste davantage malgré sa correction si les ressources cognitives sont moindres au moment d'intégrer la correction. Dans plusieurs expériences (Ecker, Lewandowsky, Swire, et al., 2011 ; Sanderson et al., 2023), des participants ont été exposés à un texte d'EIC dans lequel une charge cognitive était ajoutée au moment de la réception de la correction. Les résultats montraient que cette saturation des ressources cognitives rendait l'EIC plus important. Par ailleurs, d'autres études relatives à la mémoire de travail (Baddeley, 1992, 2010, 2012 ; Conway et al., 2005 ; Ehrlich & Delafoy, 1990) supportent également l'impact du défaut d'intégration de la correction sur l'EIC. Ce point sera développé en détail au sein de la partie 4.6.1.3. Brièvement, des éléments théoriques permettent de supposer que l'intervention

de la mémoire de travail serait nécessaire pour intégrer une correction au sein d'un modèle mental (Ecker, Oberauer, et al., 2014). Or, il est justement observé que plus un individu performe sur une tâche de mémoire de travail, moins il serait sensible à l'EIC (Brydges et al., 2018 ; Jia et al., 2020 ; McIlhiney et al., 2023).

2.2.1.3 Compétition d'activation et défaut de récupération en mémoire de la correction

En plus du moment de la réception de la correction, un second moment propice à l'EIC est lors de la compétition d'activation, lorsque le modèle mental est sollicité et doit être récupéré. A cet instant, le principe de compétition d'activation suppose que la FI et la correction devraient être simultanément activées. Ici, un EIC devrait s'observer si l'activation de la FI est plus forte que celle de la correction ou si la correction n'est simplement pas récupérée en mémoire (Ecker et al., 2010). Par exemple, une question sur les causes probables de l'incendie devrait automatiquement activer la notion de matériaux inflammables. Si le participant ne récupère pas suffisamment l'information selon laquelle le cagibi était vide, alors l'influence de la FI devrait persister.

L'hypothèse de l'effet d'une compétition d'activation entre la FI et la correction sur l'EIC est supportée par plusieurs recherches. Certaines ont observé une variation de l'EIC en manipulant le potentiel d'activation de la FI. Par exemple, Ecker et al. (2011) ont relevé un EIC plus important quand la FI avait été répétée au sein de leur texte d'EIC par rapport à si elle n'avait été présentée qu'une seule fois. Une explication possible est que la répétition de la FI l'avait rendue plus facilement accessible en mémoire, ce qui aurait augmenté son activation au moment de la récupération du modèle. Également, Ecker et al. (2015) ont exposé leurs participants à la lecture d'un scénario pour lequel deux causes complémentaires étaient proposées. L'une des causes était présentée au début de la lecture et l'autre à la fin. Leur ordre de présentation était contrebalancé et l'une des deux causes était ensuite rétractée. Les résultats

ont montré que l'EIC était plus important lorsque c'était la cause présentée à la fin de la lecture qui était rétractée par rapport à si c'était celle au début. En cohérence avec l'hypothèse d'une compétition d'activation, une interprétation des auteurs était que la dernière FI présentée était plus aisément accessible en mémoire, ce qui faciliterait son activation et donc sa persistance.

Outre l'effet de la manipulation de l'activation de la FI sur l'EIC, d'autres recherches ont montré une variation de l'EIC lorsque le potentiel d'activation de la correction était manipulé. Comme pour la FI, une manière de manipuler le potentiel d'activation de la correction peut être de la répéter puisqu'elle serait alors plus facilement accessible en mémoire au moment de la récupération. Justement, il a été observé qu'une FI persistait moins lorsque la correction avait été répétée par rapport à si elle n'avait été présentée qu'une seule fois (Ecker, Lewandowsky, Swire, et al., 2011). Par ailleurs, une autre manière de manipuler le potentiel d'activation de la correction pourrait consister à faire varier le délai entre son encodage et sa récupération. En effet, un délai plus long devrait rendre la correction moins accessible en mémoire, ce qui devrait réduire en conséquence son potentiel d'activation et donc aboutir à un EIC plus important. En accord avec cette hypothèse, il a été relevé qu'un délai de trois (Skurnik et al., 2005) ou sept jours (Ecker, Lewandowsky, et al., 2020 ; Swire et al., 2017) entre l'encodage de la correction et la récupération du modèle augmentait l'EIC par rapport à un délai immédiat.

En somme, sur la base du KReC (Kendeou et al., 2014), l'approche cognitive identifie trois situations propices à la survenue d'un EIC. Les deux premières sont au moment de l'encodage de la correction et peuvent être un défaut d'activation de la FI ou un défaut d'intégration de la correction au sein du modèle mental. La troisième est lors de la sollicitation du modèle mental et serait due à une compétition d'activation entre la FI et la correction : si la correction est insuffisamment activée pour évincer l'influence de la FI, alors celle-ci devrait s'exprimer malgré la réception de la correction.

2.2.2 Approche motivationnelle

Outre l'approche cognitive, une seconde approche explique l'EIC par des processus motivationnels (Greitemeyer, 2014 ; Johnson & Seifert, 1994 ; O'Rear & Radvansky, 2020 ; Paynter et al., 2019 ; Rich et al., 2022 ; Susmann & Wegener, 2021). Le principe général de l'approche motivationnelle est que les individus seraient motivés à rejeter la correction, ce qui mènerait par conséquent à la persistance de l'influence de la FI. A partir de la littérature, il est possible de relever deux motivations qui pourraient aboutir à un EIC.

2.2.2.1 Réticence envers l'incomplétude des modèles mentaux.

Une première explication motivationnelle de l'EIC repose, à l'instar de l'approche cognitive, sur le postulat qu'un individu exposé à un événement ou un objet en construit un modèle mental cohérent au gré des informations qu'il intègre (Johnson-Laird et al., 2012 ; Myers & O'Brien, 1998 ; O'Brien et al., 1998 ; Rapp & Kendeou, 2007). Un détail important est qu'en fonction des informations reçues, un modèle mental pourrait alors être plus ou moins complet. Par exemple, un modèle mental complet serait de connaître les causes et les conséquences d'un événement. A l'inverse, un exemple de modèle plus incomplet pourrait être d'observer un événement (e.g., l'occurrence d'un incendie) sans disposer d'explication à son occurrence (e.g., l'origine du feu). Dans cette optique, il peut être noté qu'une caractéristique des corrections est qu'elles peuvent réduire la complétude d'un modèle puisqu'elles impliquent d'en retirer une information. Par exemple, si un individu connaît les causes et les conséquences d'un incendie puis apprend que l'information sur la cause de l'incendie était finalement fausse, alors accepter la correction créerait de l'incomplétude au sein du modèle.

Un postulat complémentaire de cette approche motivationnelle est que les individus ressentent de l'inconfort face aux modèles mentaux incomplets. Cette idée est supportée par une expérience (Susmann & Wegener, 2021, étude 2) dans laquelle des participants ont lu une adaptation du texte utilisé par Johnson et Seifert (1994) puis évalué leur inconfort à la lecture

de la correction. Le texte avait toutefois été remanié afin de manipuler l'intensité de l'incomplétude laissée par la correction. Dans une condition, l'information rétractée jouait un rôle central dans l'histoire (i.e., elle portait sur la cause de l'incendie). Dans une autre, elle n'avait qu'une importance périphérique (i.e., la personne qui avait lancé l'alerte). En accord avec l'hypothèse d'un effet de l'incomplétude sur l'inconfort, les résultats ont montré que la correction d'une information centrale était jugée plus inconfortable par rapport à une FI périphérique.

A partir de ces postulats, une explication possible de l'EIC serait que la correction susciterait de l'incomplétude, donc de l'inconfort, ce qui motiverait les individus à rejeter la correction afin de réduire l'inconfort. Par exemple, dans l'expérience de Johnson et Seifert (1994), une explication de l'EIC serait que les participants n'ont pas intégré la correction (i.e., le *cagibi* était vide) car son intégration aurait impliqué l'inconfort de ne plus avoir d'explication à l'incendie. Plusieurs travaux supportent cette explication de l'EIC en montrant que la manipulation du niveau d'incomplétude laissée par la correction fait varier l'EIC. Par exemple, une manière de manipuler l'incomplétude laissée par la correction est d'augmenter l'importance de l'information rétractée au sein du modèle. Le raisonnement est que plus une information a une place centrale, plus sa correction devrait laisser un vide dans le modèle mental, et donc plus l'EIC devrait être élevé. Cette hypothèse est supportée par une recherche (Kan et al., 2021) dans laquelle des participants ont lu un texte d'EIC. Le caractère central de la FI était manipulé entre deux groupes de participants, c'est-à-dire qu'elle avait un rôle central pour un groupe et périphérique dans un autre. En accord avec l'hypothèse d'un impact de l'incomplétude sur l'EIC, la FI persistait davantage malgré sa correction lorsqu'elle était centrale. Par ailleurs, une autre manière de manipuler l'incomplétude serait d'ajouter, à la correction, une explication alternative à la FI plutôt que de simplement dire qu'elle est fautive. En effet, l'explication alternative permettrait de combler la lacune que laisserait la suppression de la FI, ce qui devrait

réduire l'inconfort et donc l'EIC. Cette hypothèse a été testée par plusieurs expériences. Par exemple, dans une recherche de Johnson et Seifert (1994), la correction était une simple rétractation pour un groupe (e.g., « *le cagibi était vide* ») et contenait une explication alternative pour un autre (e.g., « *le cagibi était vide, mais plusieurs traces laissent penser à un incendie criminel* »). Dans cette recherche et dans de nombreuses autres (Ecker et al., 2023 ; Hamby et al., 2020 ; Jin et al., 2022 ; Johnson & Seifert, 1994 ; Kan et al., 2021 ; Kendeou et al., 2014 ; MacFarlane et al., 2018 ; Paynter et al., 2019 ; Rapp & Kendeou, 2007), l'ajout d'une explication alternative permettait de réduire l'EIC. Ainsi, un premier argument à l'approche motivationnelle basée sur l'incomplétude est que le vide laissé par la correction au sein du modèle semble prédire la persistance de la FI.

Un second argument à l'approche motivationnelle basée sur l'incomplétude est qu'une réévaluation de l'incomplétude au moment de la correction influence l'EIC, ce qui est supporté par une expérience de Susmann et Wegener (2021, étude 3). Avant leur lecture d'un texte d'EIC, un groupe de participants a reçu une série d'instructions leur suggérant qu'il était normal de ressentir de l'inconfort face à des événements dont les justifications sont incomplètes. Les résultats ont démontré que cette manipulation de l'évaluation de l'inconfort permettait une réduction de l'EIC. En somme, en montrant que cet effet varie lorsque l'inconfort ou le vide suscité par la correction sont manipulés, ces différents travaux démontrent qu'une explication possible de l'EIC réside dans une réticence à accepter un modèle mental incomplet.

2.2.2.2 Raisonement motivé

Une explication motivationnelle complémentaire de l'EIC est que les individus pourraient rejeter ou ignorer la correction si elle va à l'encontre de leurs attitudes initiales. La notion d'« assimilation biaisée » désigne ce phénomène au sein duquel une même information est évaluée différemment en fonction des attitudes des individus (Lord et al., 1979 ; McHoskey, 1995). Par exemple, un article scientifique en faveur de la peine de mort est jugé plus

convaincant et de meilleure qualité par les individus favorables envers cette sentence par rapport aux autres (Lord et al., 1979). De nombreuses études démontrent également que des informations qui dégradent l'image d'un candidat ou d'un parti politique suscitent davantage d'adhésion chez les participants qui ont une idéologie politique opposée au candidat et au parti incriminés, comparativement à ceux qui le supportent (Dobber et al., 2021 ; Garrett, 2011 ; Pennycook et al., 2018 ; Prior et al., 2015 ; Weeks & Garrett, 2014 ; Winneg et al., 2005).

Une explication à cette assimilation biaisée des informations réside dans la littérature autour du raisonnement motivé (Kahan, 2013 ; Kunda, 1987, 1990) qui propose que l'évaluation des informations peut être guidée par deux motivations distinctes. Dans certaines situations, l'évaluation des informations se porterait sur l'exactitude, c'est-à-dire que les individus chercheraient la précision et l'objectivité dans leur jugement. Par exemple, un individu pourrait évaluer les informations sur l'invasion de l'Irak par les Etats-Unis avec comme objectif de savoir réellement ce qui s'y est produit. Dans d'autres situations, le raisonnement pourrait être motivé par la volonté d'arriver à un résultat particulier, qu'il soit exact ou non. Un individu pourrait par exemple évaluer les informations sur l'invasion de l'Irak par les Etats-Unis avec comme motivation d'aboutir à la conclusion que le gouvernement américain avait raison lorsqu'il affirmait que le pays détenait des armes de destruction massive. Cette dernière motivation serait à la source d'une assimilation biaisée puisque les informations sont alors traitées différemment en fonction des attitudes des individus. Selon les auteurs, cette motivation serait notamment activée lorsque le contenu de l'information est important pour l'individu.

Dans le cadre de l'EIC, la conséquence du raisonnement motivé par la volonté d'arriver à un résultat particulier est que les individus traiteraient différemment les corrections en fonction de leurs attitudes. En l'occurrence, si une correction est contre-attitudinale, alors les individus pourraient être motivés à la rejeter. Plusieurs recherches supportent cette

hypothèse en montrant une variation de l'EIC en fonction du caractère pro- ou contre-attitudinale de la correction. Par exemple dans une expérience (Nyhan & Reifler, 2010), des participants américains ont été exposés à une phrase du président conservateur Bush suggérant que l'Irak aurait détenu des armes de destruction massive. Un rapport réfutait ensuite cette FI puis l'adhésion des participants à cette dernière et leur idéologie politique étaient évaluées. Les résultats montraient alors un effet d'interaction entre l'idéologie politique et l'efficacité corrective du rapport : la correction réduisait l'adhésion à la FI pour les participants démocrates mais pas pour les conservateurs (voir aussi Amazeen et al., 2018 ; Ecker & Ang, 2019 ; Nyhan et al., 2020). De la même manière aux Etats-Unis, les individus conservateurs auraient moins tendance à adhérer au consensus sur l'origine anthropogène du changement climatique (Dunlap & McCright, 2008). En accord avec l'hypothèse motivationnelle basée sur le raisonnement motivé, des recherches ont observé qu'un message rappelant ce consensus était moins efficace chez les conservateurs par rapport aux démocrates (Bolsen & Druckman, 2018 ; Cook & Lewandowsky, 2016). Les travaux autour de l'interaction entre les attitudes initiales et l'EIC seront davantage développés au sein de la partie 4.3.5.

Pour résumer, la littérature propose deux explications motivationnelles à l'EIC. L'une est basée sur le fait que l'incomplétude d'un modèle mental susciterait de l'inconfort, ce qui motiverait les individus à rejeter la correction. L'autre est basée sur le raisonnement motivé et propose que les individus seraient motivés à rejeter une correction si celle-ci est contre-attitudinale, ce qui aboutirait à la persistance de la FI malgré sa correction. Il doit toutefois être souligné que la portée de cette dernière explication motivationnelle se limite aux situations dans lesquelles les attitudes initiales sont contredites.

2.2.3 Approche rationnelle

Une dernière approche de l'EIC consiste à expliquer cet effet par une compétition de crédibilité entre la correction et la FI. L'hypothèse générale est que l'EIC s'exprimerait lorsque

les individus croient davantage à la FI qu'en la correction. Par exemple, si un individu a pour principe de ne jamais croire ce qu'il voit dans les médias, alors il serait attendu qu'il n'adhère pas à la correction selon laquelle l'Irak ne détenait pas d'armes de destructions massives si la source de cette correction est un journal, et qu'il continue par conséquent de croire en la FI. Cette approche est dite « rationnelle » (Connor Desai et al., 2020) car il est rationnellement attendu que les cognitions s'alignent avec ce que l'individu croit et s'éloignent de ce qu'il trouve douteux.

L'explication rationnelle est supportée par les recherches qui montrent une variabilité de l'EIC en manipulant la crédibilité de la correction. Par exemple, des expériences ont exposé des sujets à une correction dont la source était faiblement crédible pour un groupe et fortement crédible pour un autre. Les résultats montraient que la FI persistait moins lorsque la correction était portée par une source plus crédible (Guillory & Geraci, 2013 ; voir aussi Ecker & Antonio, 2021). Également, sur les réseaux sociaux, la correction d'une fausse croyance sur une maladie était plus efficace lorsqu'elle était donnée par une institution de santé par rapport à un simple utilisateur (Vraga & Bode, 2017b), ou encore lorsqu'un utilisateur faisait référence à une institution de santé dans sa correction (Vraga & Bode, 2017a). Dans une autre recherche, des sujets ont lu des textes d'EIC. Pendant qu'ils lisaient ces textes et avant de répondre à des questions inférentielles, ils ont eu pour consigne d'évaluer la crédibilité de chaque message. Les résultats montraient que la crédibilité attribuée à la correction prédisait le nombre d'inférences liées à la FI : moins la correction suscitait d'adhésion, plus la FI persistait (Ecker & Antonio, 2021). D'autres évidences concernant l'effet de la crédibilité de la correction seront apportées au sein de la partie 4.3.3.

En plus de ces études manipulant la crédibilité de la correction, il existe d'autres recherches qui supportent l'approche rationnelle en manipulant la crédibilité de la FI, bien que ces évidences restent indirectes. Par exemple, une expérience a exposé un groupe de

participants à un texte d'EIC dans lequel la FI avait été validée comme étant atypique (e.g., l'incendie aurait été causé par les débris d'un satellite) tandis qu'un autre recevait une FI évaluée comme étant typique (e.g., l'incendie aurait été causé par des bonbonnes de gaz). Après la correction, les résultats montraient que la persistance de la FI était plus importante lorsque la FI était typique (Rich et al., 2022, étude 2). En accord avec l'approche rationnelle, une explication possible est que la FI typique était évaluée comme plus crédible que la FI atypique, favorisant ainsi sa crédibilité dans sa compétition avec la correction. Une autre évidence indirecte est portée par Ecker et al. (2011), qui ont observé qu'une FI persistait davantage si elle avait été répétée par rapport à si elle n'avait été présentée qu'une fois. Toute la littérature autour de l'effet de véracité (Dechêne et al., 2010) montre que la répétition d'une information la rend plus crédible (Fenn et al., 2013 ; Pennycook et al., 2018 ; Swire et al., 2017 ; Weaver et al., 2007). Par exemple, une recherche observait qu'une proposition ambiguë telle que « *Canopus est le nom de l'étoile la plus brillante dans le ciel, exception faite du soleil.* » était plus souvent jugée vraie si elle avait été répétée par rapport à si elle n'était présentée qu'une seule fois (Pennycook et al., 2018). Par conséquent, le constat que l'EIC augmente lorsque la FI est répétée (Ecker et al., 2011) peut indirectement supporter l'approche rationnelle. L'effet de la crédibilité de la FI sur l'EIC sera également discuté au sein de la partie 4.1.2.

Ainsi, l'approche rationnelle propose d'expliquer l'EIC par une compétition de crédibilité entre la FI et la correction. Elle est supportée par les recherches qui montrent une variabilité de cet effet lorsque la crédibilité de ces informations est manipulée. Pour finir, il peut être relevé ici que l'approche rationnelle reste relativement peu citée dans la littérature sur l'EIC comparativement aux explications cognitives et motivationnelles.

2.3 Autres possibilités à la suite de la correction d'une fausse information

En plus de l'EIC, il est important de noter que d'autres issues sont envisageables à la suite de la correction d'une FI. Elles sont un effet boomerang, un effet de sur-correction, ou encore un retour ou un maintien de la ligne de base.

2.3.1 Effet boomerang

Une première possibilité à la suite de la correction d'une FI est un effet boomerang. De façon générale, un effet boomerang désigne un changement des cognitions dans le sens contraire à celui qu'un stimulus était censé induire (Hart & Nisbet, 2012). Cet effet peut aussi prendre le nom d'effet « backfire » (i.e., retour de flamme), « backlash » (i.e., retour de manivelle) (Cook & Lewandowsky, 2012 ; Nyhan et al., 2014) ou plus rarement d'« effet cobra ». Cette dernière appellation fait référence à un évènement qui se serait déroulé lors de la période coloniale en Inde dans la ville de Delhi. Pour lutter contre la prolifération de cobras, le gouvernement aurait proposé une prime aux citoyens qui rapporteraient des dépouilles de serpents. En réponse, certains habitants auraient développé des élevages de reptiles dans le but de percevoir la bourse. Face à cette situation, le gouvernement aurait alors annulé la prime, ce qui aurait finalement motivé les éleveurs à relâcher leurs cobras dans la ville (Bronner, 2021, p. 76).

2.3.1.1 Mise en évidence de l'effet boomerang

Depuis, un effet boomerang a été démontré dans de nombreux contextes. Par exemple, un message censé augmenter les donations pour des enfants dans le besoin peut avoir l'effet inverse de réduire les dons (Small et al., 2007), un message de prévention contre le tabac augmenter le souhait de fumer (Wolburg, 2006), un message visant à réduire la consommation d'énergie l'accroître chez certains ménages (Schultz et al., 2007), des programmes cherchant à décroître la consommation d'alcool d'adolescents l'augmenter (Wechsler et al., 2003) et l'avertissement de la présence de contenus violents dans un film élever le désir de le voir

(Bushman & Stack, 1996). Divers effets boomerangs ont aussi été observés dans la littérature relative aux FI. Par exemple, une expérience a relevé que l'exposition d'internautes à des corrections pouvait réduire la qualité des informations qu'ils partageaient ensuite (Mosleh et al., 2021). De même, dans une autre recherche conduite sur les réseaux sociaux, il était observé que des internautes qui aimaient les contenus de pages complotistes consultaient davantage ces contenus après avoir été exposés à un élément de *fact-checking* (Zollo et al., 2017). Enfin, d'autres travaux ont exposé des participants à des articles scientifiques fictifs contredisant leurs croyances et a observé que cette manipulation réduisait la confiance des participants envers la science ainsi que leur propension à utiliser la science pour résoudre des problèmes (Drummond et al., 2020 ; Munro, 2010).

Dans le cadre de la correction des FI, l'application d'un effet boomerang se traduirait par le fait que la correction d'une FI conduise à des cognitions s'alignant davantage dans le sens de la FI par rapport à si la correction n'avait pas été présentée (Guess & Coppock, 2020). Cette issue est illustrée dans la Figure 2. Plusieurs études ont illustré un effet boomerang à la suite de la correction d'une FI (voir cependant partie 2.3.1.3). Par exemple, dans le domaine militaire ou politique, des participants ont lu le discours d'un président américain conservateur suggérant la présence d'armes de destruction massive en Irak, qui était ensuite réfuté. Un effet boomerang était alors observé chez certains participants qui devenaient *plus* enclins à croire qu'il y avait des armes de destruction massive en Irak par rapport à d'autres n'ayant pas lu la correction (Nyhan & Reifler, 2010, étude 1). De la même manière, d'autres recherches ont observé que la correction d'une affirmation qui favorisait le parti républicain (Nyhan & Reifler, 2010, étude 2) ou d'une information qui incriminait le parti démocrate (Nyhan et al., 2013) augmentait l'adhésion à ces FI chez certains participants. Dans un autre contexte, des travaux ont relevé que des messages scientifiques contredisant le lien entre le développement de l'autisme et le vaccin ROR réduisaient la propension de parents à vacciner leurs enfants (Nyhan

et al., 2014). Par exemple, dans une expérience (Pluviano et al., 2017), des participants ont été exposés à des messages rappelant les mythes et les faits autour de la vaccination. Une semaine après l'intervention (mais pas immédiatement après), cette communication *augmentait* l'hésitation vaccinale ainsi que la croyance en l'impact causal de la vaccination sur l'autisme et en des effets secondaires non-avérés (voir aussi Nyhan & Reifler, 2015). Pour finir, une autre étude a observé que l'exposition à un message explicitant les dangers du réchauffement climatique réduisait le support envers les politiques de lutte contre le réchauffement climatique pour certains participants (Hart & Nisbet, 2012).

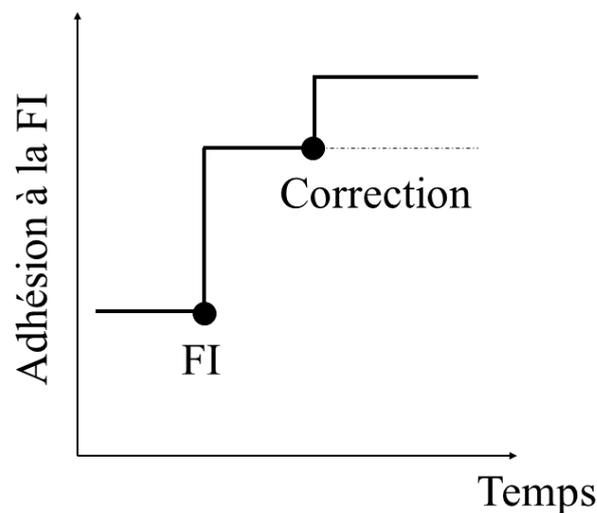


Figure 2. Illustration d'un effet boomerang à la suite de la correction d'une FI

2.3.1.2 Explication théorique

La littérature identifie deux processus qui pourraient expliquer un effet boomerang à la suite de la correction d'une FI (Swire-Thompson et al., 2020). Le premier dépend de la vision du monde de l'individu et le second d'un effet de familiarité de la FI.

2.3.1.2.1 Effet boomerang dû à la vision du monde

Un premier processus qui pourrait susciter un effet boomerang est lorsque la correction de la FI contredit la vision du monde de l'individu (*worldview backfire effect*, Lewandowsky et al., 2012). Par exemple, dans le domaine politique aux Etats-Unis, il est attendu que des individus en faveur du parti républicain adhèrent à des informations pro-conservatrices, anti-

démocrates, et aient plus tendance à renier la véracité et l'importance du réchauffement climatique (Dunlap & McCright, 2008). Justement, des recherches ont exposé des individus à des corrections de FI pro-conservatrices (Nyhan & Reifler, 2010), anti-démocrates (Nyhan et al., 2013) ou de FI reniant les dangers du réchauffement climatique (Hart & Nisbet, 2012). Supportant l'hypothèse d'une interaction entre la vision du monde et l'effet boomerang, les résultats ont observé une hausse de l'adhésion à ces FI après leur correction chez les participants républicains mais pas chez les démocrates. Dans un domaine différent, des expériences avaient observé un effet boomerang suite à la présentation de messages spécifiant l'innocuité des vaccins ROR (Nyhan et al., 2014) ou contre la grippe (Nyhan & Reifler, 2015). Ici aussi, l'effet boomerang relevé ne s'observait que pour les participants dont la vision du monde semblait s'opposer aux messages correctifs, tels que des individus ayant préalablement manifesté des préoccupations par rapport aux effets secondaires (Nyhan & Reifler, 2015) ou des attitudes défavorables envers la vaccination (Nyhan et al., 2014).

Une explication possible à l'interaction entre la vision du monde et l'effet boomerang est qu'une correction contre-attitudinale stimulerait l'élaboration de contre-arguments (Sherman & Cohen, 2002), ce qui pourrait susciter un effet boomerang par deux moyens. Le premier est que, face à la correction, les individus récupérerait en mémoire des contre-arguments déjà existants. Ce processus pourrait alors augmenter l'adhésion à la FI car ils auraient plus d'éléments activés la supportant. Un deuxième processus possible est que l'exposition à la correction motive les individus à rechercher des contre-arguments et que par conséquent, ils en développent de nouveaux. Ces explications sont toutefois spéculatives. A notre connaissance, les effets médiateurs de l'activation et la création de contre-arguments sur l'effet boomerang doivent encore être démontrés.

2.3.1.2.2 *Effet boomerang dû à la familiarité*

Le second processus par lequel une correction pourrait susciter une hausse de l'adhésion à la FI est l'effet de familiarité (*familiarity backfire effect*, Lewandowsky et al., 2012), bien que ce processus se limite aux corrections qui mentionnent explicitement la FI (e.g., « *Le cagibi qui était supposé contenir des matériaux inflammables était vide.* ») Le postulat de cette explication est que les individus utilisent, en plus de leurs connaissances, plusieurs heuristiques pour juger de la véracité d'une proposition (Brashier & Marsh, 2020). Parmi ces raccourcis mentaux, l'un d'eux serait la fluidité avec laquelle l'information est traitée. Concrètement, si elle est traitée de façon fluide et avec facilité, elle aurait plus tendance à être perçue comme vraie par rapport à si elle était traitée difficilement (Koch & Forgas, 2012). Or, une manière d'augmenter la fluidité dans le traitement d'une information serait de la répéter (Marsh & Stanley, 2021). Pour cause, une information devrait être traitée plus facilement si elle a déjà été traitée au préalable. La conséquence est qu'une information qui a été répétée devrait paraître plus vraisemblable par rapport à si elle ne l'a pas été (Dechêne et al., 2010 ; Fenn et al., 2013 ; Jin et al., 2022 ; Pennycook et al., 2018 ; Swire et al., 2017 ; Weaver et al., 2007).

L'hypothèse de l'effet boomerang dû à un effet de familiarité est qu'une correction qui mentionne explicitement la FI pourrait augmenter sa familiarité, donc son caractère vraisemblable, et donc l'adhésion qu'elle suscite. Par exemple, la correction d'une FI relative à la dangerosité d'un médicament, telle que « *Il n'est pas vrai que le médicament est dangereux.* » aurait comme conséquence de répéter l'association entre « médicament » et « danger » et pourrait donc faire paraître la FI plus vraie par rapport à si elle n'avait pas été corrigée. Une expérience qui supporte cette hypothèse est celle de Skurnik et al. (2005). Dans une première partie de cette recherche, des participants ont été exposés à plusieurs affirmations tout en leur révélant, à chaque présentation, si elles étaient vraies ou fausses. Certaines affirmations et leur correction étaient présentées trois fois et d'autres seulement une. Dans une deuxième étape,

après un délai de trois jours, les participants ont eu pour consigne d'évaluer la véracité des affirmations précédemment présentées. Un effet boomerang a alors été observé chez les participants âgés (mais pas les plus jeunes), qui avaient plus tendance à déclarer véridiques les FI qui avaient été répétées et corrigées par rapport à celles qui ne l'avaient pas été. Un autre travail qui, selon les auteurs, supporte l'hypothèse d'un effet boomerang dû à la familiarité est celui de Pluviano et al. (2017). Les auteurs ont observé que des messages opposant les mythes et les faits autour de la vaccination augmentait l'adhésion des participants à des fausses croyances une semaine après l'intervention (mais pas immédiatement). Leur interprétation était que la répétition des mythes lors de l'intervention avait comme conséquence de les rendre plus familiers, augmentant alors leur caractère vraisemblable. Cet effet boomerang dû à un effet de familiarité a amené plusieurs auteurs à préconiser de ne pas mentionner explicitement les FI au sein des corrections (Cook & Lewandowsky, 2012 ; Marsh & Stanley, 2021).

2.3.1.3 Remise en question de l'effet boomerang suite à la correction d'une fausse information

Malgré les quelques évidences autour de l'effet boomerang et les inquiétudes qu'il a suscitées, il est important de noter que beaucoup de travaux remettent en question sa validité, qu'il soit dû à un effet de familiarité ou à une interaction avec la vision du monde.

2.3.1.3.1 *Remise en question de l'effet boomerang dû à un effet de familiarité*

Concernant l'effet boomerang dû à un effet de familiarité, plusieurs expériences ont observé une absence de répliation de ce phénomène, voire un effet inverse. Par exemple, Ecker et al. (2023) ont échoué dans leur tentative de répliation de l'étude de Pluviano et al. (2017) qui avait observé un effet boomerang après l'exposition à des messages opposant des mythes aux faits sur la vaccination. De la même manière, Ecker et al. (2020) ont comparé l'adhésion de participants à des FI après qu'elles aient été corrigées en étant mentionnées explicitement (e.g., « *Aux Etats-Unis en 2016, 79 % des victimes de couleur de peau blanche ont été tuées*

par des individus avec une couleur de peau noire. FAUX») par rapport à leur adhésion avant leur rétractation et par rapport à l'adhésion d'un groupe contrôle n'ayant pas été exposé à la correction. Quelle que soit la ligne de base utilisée, aucun effet boomerang n'a été observé malgré la répétition de la FI au sein de la correction (voir aussi Swire et al., 2017). Enfin, dans une autre expérience (Sanderson et al., 2023), il était observé que la répétition d'une rétractation (e.g., « *Notre hypothèse initiale selon laquelle l'incendie avait été déclenché volontairement était infondée [...] nous pensons maintenant que l'incendie n'est pas de source criminelle [...], les gens doivent être rassurés de savoir que l'incendie n'était pas de source criminelle* ») n'engendrait pas d'effet boomerang par rapport à des participants pour lesquels la correction n'était pas répétée (e.g., « *Après de plus amples investigations, nous pensons finalement que l'incendie n'est pas de source criminelle.* ») ou pas présentée. Au contraire, il était observé que la répétition de la correction était plus efficace par rapport à si elle n'était présentée qu'une seule fois (voir aussi Ecker et al., 2011).

Il doit être souligné que l'absence d'effet boomerang dû à un effet de familiarité s'observe même dans des conditions qui lui seraient pourtant propices. Une situation propice serait notamment de répéter la FI tout en complexifiant la récupération de la correction, car cette situation devrait orienter la compétition d'activation en faveur de la FI. Suivant ce raisonnement, des expériences (Ecker et al., 2023 ; Ecker, Lewandowsky, et al., 2020 ; Sanderson et al., 2023) ont examiné si cet effet s'observait lorsqu'une FI avait été répétée et que la récupération de la correction était entravée par différentes variables telles que l'ajout d'une charge cognitive au moment de son encodage (Sanderson et al., 2023), l'attente d'un délai d'une ou trois semaines pour solliciter le modèle mental, ou encore en évaluant des participants âgés de plus de 50 ans (Ecker et al., 2023 ; Ecker, Lewandowsky, et al., 2020), censés avoir de moins bonnes capacités pour récupérer des informations en mémoire (Grady &

Craik, 2000). Malgré leur caractère conservateur envers l'hypothèse d'un effet boomerang, ces études ne l'ont pas confirmée.

La seule situation qui continue de susciter des inquiétudes quant à un potentiel effet boomerang dû à un effet de familiarité est lorsqu'une correction reprend une FI qui n'avait jamais été présentée auparavant (Autry & Duarte, 2021 ; Ecker, Lewandowsky, et al., 2020). En ce sens, une expérience (Autry & Duarte, 2021) a exposé des participants à plusieurs textes. Certains comportaient la correction explicite d'une information qui n'avait pas été présentée auparavant tandis que d'autres ne comportait pas de FI ni de correction. Par exemple, dans un texte décrivant un cambriolage, certains participants apprenaient qu'une voiture suspecte n'était pas bleue, sans qu'il ne leur ait été fait mention de la couleur de la voiture auparavant. Pour d'autres participants, aucune mention de la couleur de la voiture n'était produite. Les inférences des individus sur les textes étaient ensuite interrogées. Dans deux expériences sur ce paradigme, utilisant six textes différents, les résultats ont mis en avant un effet boomerang : les participants exposés à la négation d'une information nouvelle faisaient davantage référence à l'information corrigée dans leurs réponses aux questions inférentielles par rapport à des participants nullement exposés à l'information. De la même manière, Ecker et al. (2020) ont exposé un groupe de participants à des corrections d'affirmations validées comme étant inconnues (e.g., « *FAUX : Facebook s'apprête à lancer une campagne contre les grossièretés.* ») Une semaine après leur exposition à ces messages, l'adhésion des participants aux informations corrigées était comparée à des individus n'y ayant pas été exposés. Les résultats de leur première expérience montraient un effet boomerang de faible taille (Ecker et al., 2020, étude 1), c'est-à-dire que les participants exposés à la correction de l'affirmation inconnue y adhéraient davantage par rapport à ceux qui n'y avaient pas été exposés. Il doit toutefois être noté que Ecker et al. (2020) ont mené une tentative de réplique de leurs résultats dans deux expériences différentes en utilisant un échantillon plus important et que celle-ci a échoué (Ecker et al., 2020,

étude 2, étude 3). Leurs résultats allaient à l'inverse dans le sens d'une efficacité corrective des messages. Face à ces évidences contradictoires, les recommandations actuelles restent d'éviter de corriger une FI en la répétant si elle n'a pas encore été répandue. Dans les autres situations où la FI est déjà diffusée, l'absence de réplication de l'effet boomerang dû à un effet de familiarité amène à remettre en question la validité de cet effet (Ecker et al., 2022 ; Sanderson et al., 2023 ; Swire-Thompson et al., 2020).

2.3.1.3.2 Remise en question de l'effet boomerang dû à la vision du monde

Il est aussi important de souligner que de nombreux travaux n'observent pas de réplication de l'effet boomerang dû à la vision du monde, même lorsque les conditions lui sont propices. Par exemple, une recherche (Ecker et al., 2023) a réévalué l'occurrence d'un effet boomerang suite à des messages corrigeant des fausses croyances sur la vaccination (Nyhan et al., 2014 ; Nyhan & Reifler, 2015) et n'a pas répliqué cet effet. De la même manière, comme mentionné précédemment (voir partie 2.3.1.2.1), des travaux ont noté que la correction de FI pro-attitudinales chez des individus républicains (pro-conservatrices ou anti-démocrates) suscitait une hausse de l'adhésion à ces FI (Nyhan et al., 2013 ; Nyhan & Reifler, 2010). Or, il est pertinent de noter que si cet effet s'observait chez les individus républicains, il ne s'observait pas chez les individus démocrates lorsqu'une FI pro-démocrate ou anti-conservatrice était réfutée (Nyhan & Reifler, 2010), ce qui interroge la prévalence de cet effet. De plus, cet effet boomerang n'a pas été répliqué dans l'étude d'Amazeen et al. (2018) qui exposait des individus à la correction d'une information déclarée par une personnalité affiliée à leur parti politique (voir aussi Weeks & Garrett, 2014). Enfin, Wood et Porter (2016) ont mené quatre expériences, au sein desquelles plus de 8000 participants ont été recrutés. Un total de 36 affirmations politiques fausses, allant dans le sens de diverses orientations politiques ont été présentées aux participants avant d'être réfutées. Parmi toutes ces propositions, une seule concernant les armes de destruction massive en Irak a provoqué un effet boomerang chez les individus conservateurs

et cet effet disparaissait lorsque l'item utilisé était modifié. Il peut être noté que cette absence d'effet boomerang s'observait alors même que la recherche était conduite au moment des primaires des élections américaines de 2016, une période probablement propice à l'activation et l'élaboration de contre-arguments. En somme, tous ces travaux amènent à remettre en question la prévalence de l'effet boomerang à la suite de la correction d'une FI, qu'il soit dû à la vision du monde ou à la familiarité. Le consensus actuel semble que l'effet boomerang à la suite de la correction d'une FI n'est pas un effet fantasmatique, mais que son occurrence reste très rare et ne devrait donc pas être une raison pour empêcher des tentatives de corriger des FI (Guess & Coppock, 2020 ; Swire-Thompson et al., 2020 ; Weeks & Garrett, 2014 ; Wood & Porter, 2016).

2.3.2 *Sur-correction*

Outre l'effet boomerang, un autre effet qui pourrait survenir à la suite de la correction d'une FI est un effet de sur-correction. Le principe général de cet effet est que lorsque les individus identifient qu'un biais influence leur jugement, ils pourraient chercher à compenser ce biais en renversant leur jugement dans le sens opposé. Si cette compensation est toutefois excessive, ce processus aboutirait à la formation d'une évaluation biaisée dans le sens opposé au biais initial.

2.3.2.1 Mise en évidence de l'effet de sur-correction

De manière générale, l'effet de sur-correction a été inféré dans plusieurs travaux qui ont observé que, dans certaines situations censées biaiser une attitude dans une certaine direction, l'attitude se polarisait excessivement dans le sens inverse du biais. Par exemple, une expérience (Wegener & Petty, 1995, étude 3) a développé un paradigme propice à un effet de contraste, c'est-à-dire à un contexte qui biaiserait l'évaluation d'un stimulus dans le sens opposé au contexte. Des participants ont évalué l'attractivité d'individus avant et après avoir été exposés à des actrices attirantes. Un effet de contraste était observé pour la plupart des

participants, qui jugeaient moins attractifs les individus après l'exposition aux actrices. Cependant, lorsque les participants recevaient la consigne de ne pas laisser leur évaluation être influencée par la vision préalable des actrices, un effet de sur-correction semblait survenir. A la suite de l'exposition aux actrices, l'évaluation des individus devenait *plus positive* par rapport à l'évaluation initiale. De la même manière, d'autres recherches ont développé des paradigmes propices à des effets d'assimilation, c'est-à-dire à des contextes qui biaiserait l'évaluation d'un stimulus dans le même sens que le contexte. Un effet d'assimilation consiste par exemple à juger *plus probables* des événements tristes suite à une induction émotionnelle de tristesse, ce qui a été démontré par plusieurs expériences (DeSteno et al., 2000). Néanmoins, ces mêmes expériences ont inféré un effet de sur-correction chez certains participants qui jugeaient *moins probables* l'occurrence d'évènements tristes à la suite d'une induction de tristesse. Dans un autre contexte, une recherche (Lambert et al., 1997) a invité des participants à évaluer les qualités d'une personne appartenant ou non à un groupe associé à des stéréotypes négatifs. A la suite d'une induction d'état émotionnel positif ou sans manipulation de l'état émotionnel, les résultats montraient une moins bonne évaluation de la personne si elle appartenait au groupe stigmatisé, démontrant l'influence du stéréotype. Cependant, un effet de sur-correction était inféré lorsque l'émotion induite était de la tristesse. Dans cette situation, l'individu était évalué de façon *plus positive* lorsqu'il appartenait au groupe associé à des stéréotypes négatifs par rapport à s'il n'y appartenait pas. Enfin, le même impact de l'induction de tristesse était relevé dans une autre expérience s'intéressant à l'heuristique selon lequel ce qui est beau est bon (Amadiou, 2002). A la suite d'une induction de joie ou sans manipulation de l'état émotionnel, des sujets avaient plus tendance à recruter une personne si elle était attractive par rapport à si elle ne l'était pas, confirmant l'utilisation générale de l'heuristique. Un effet de sur-correction était toutefois inféré suite à une induction émotionnelle de tristesse, où les participants avaient

moins tendance à recruter la personne attractive par rapport à si l'humeur n'avait pas été manipulée (Lambert et al., 1997).

2.3.2.2 Explication théorique

Un cadre théorique important autour de l'effet de sur-correction est le *Flexible Correction Model* (FCM, voir Wegener & Petty, 1997). Le principe de ce modèle est que l'effet de sur-correction surviendrait lorsqu'un individu identifie que son jugement est potentiellement biaisé, et compense alors son jugement dans le sens inverse du biais afin de le corriger. Par exemple, dans le cadre de l'expérience de DeSteno et al. (2000), une explication serait que les participants montrant un effet inverse au biais attendu ont détecté que leur état émotionnel triste pourrait augmenter leur estimation de la probabilité d'évènements tristes, les amenant à revoir à la baisse leur estimation initiale.

A partir de la littérature sur le FCM (Wegener & Petty, 1995, 1997), nous proposons de synthétiser ce modèle à partir de cinq principes. Un premier est que les évaluations des individus peuvent être biaisées par des éléments de l'environnement. Ce point a été supporté par les différentes expériences montrant l'influence d'effets de contraste (Wegener & Petty, 1995), d'assimilation (DeSteno et al., 2000), d'un stéréotype ou encore d'un heuristique (Lambert et al., 1997) sur l'évaluation d'individus.

Un second postulat du FCM est que les individus disposent de théories sur ces biais, c'est-à-dire qu'ils sont capables de juger, à tort ou à raison, lorsque leur jugement serait ou non influencé. En ce sens, une expérience a demandé à des participants s'ils pensaient que la présentation d'un certain contexte pourrait impacter leur évaluation d'un stimulus cible. Les résultats ont montré qu'ils avaient effectivement une théorie sur les effets de contraste et d'assimilation. Par exemple, ils répondaient explicitement que la vue d'actrices attirantes pourrait rendre leur évaluation du physique d'une personne d'apparence normale plus négative (effet de contraste) et répondaient que l'évaluation d'un produit pourrait être rendue plus

positive s'il était approuvé par ces actrices (effet d'assimilation) (Wegener & Petty, 1995, étude 1).

Un troisième postulat est que les individus seraient capables de corriger ces biais par un procédé d'auto-correction. Une première évidence à ce sujet est portée par le travail de Wegener et Petty (1995) qui observait une réduction des effets de contraste et d'assimilation lorsque les participants étaient incités à ne pas laisser leur jugement être influencé par le contexte. Une autre illustration possible de l'auto-correction est amenée par une expérience de Kubovy (1977). Dans cette recherche, des participants ont été invités à indiquer le premier chiffre aléatoire qui leur venait à l'esprit. Lorsque la consigne était simplement d'indiquer un chiffre au hasard, le nombre 1 était choisi dans 18 % des cas. Toutefois, si la consigne était d'indiquer un chiffre au hasard *tel que 1*, alors le nombre 1 n'était choisi que dans 5,4 % des cas. En accord avec l'idée d'auto-correction, l'auteur supposait que dans la deuxième condition, les participants avaient pris conscience que l'exemple du 1 pourrait influencer leur réponse, que cette influence ne correspondait plus à leur idée du hasard, et qu'ils se seraient alors auto-corrigés en évitant ce chiffre.

Un quatrième principe du FCM est que l'auto-correction est une action réalisable sous deux conditions. Une première est que l'individu soit motivé à avoir un jugement sans biais. En ce sens, des recherches ont observé que des biais s'exprimaient moins chez des personnes qui déclaraient être motivées par le fait de s'engager dans des activités cognitives (DeSteno et al., 2000 ; Martin et al., 1990, étude 3), ce qui peut être un indicateur de leur motivation à avoir un jugement sans biais. De même, une autre expérience a placé (Martin et al., 1990, étude 2) des participants dans une situation propice à un effet d'assimilation et a observé que cet effet disparaissait lorsqu'il était indiqué aux participants que leurs réponses seraient jugées par des pairs, ce qui peut être expliqué par une motivation accrue à éviter les biais. Une deuxième condition à l'auto-correction est la disponibilité de ressources cognitives suffisantes. Cette

proposition est supportée par une expérience (Martin et al., 1990, étude 1) dans laquelle des participants ont été placés dans une situation propice à un biais d'assimilation. Les participants montraient effectivement un biais d'assimilation lorsqu'ils réalisaient simultanément une tâche distractive, mais ne le montraient plus lorsque la tâche n'était pas demandée.

Enfin, un dernier postulat du FCM est que le degré d'auto-correction d'un biais serait proportionnel à l'influence perçue de ce biais. Supportant cette hypothèse, une expérience a interrogé des participants sur leurs théories envers les effets de contraste et d'assimilation. Les sujets étaient ensuite placés dans des situations propices à l'auto-correction de ces biais (i.e., ils recevaient la consigne explicite de ne pas se faire influencer). Confirmant le postulat, les résultats ont montré que plus les participants estimaient l'influence des biais importante, moins ces biais s'exprimaient dans les réponses (Wegener & Petty, 1995, étude 4).

Un point notable relatif au dernier postulat est que les individus révisent leur jugement en fonction du biais qu'ils perçoivent et non pas du biais réel. Or, il est important de noter que le biais réel peut diverger par rapport au biais perçu : il peut être moins important, voire ne pas exister ou aller dans la direction opposée. Selon le FCM (Wegener & Petty, 1995, 1997), c'est précisément cette divergence entre le biais réel et le biais perçu qui expliquerait l'effet de sur-correction. En somme, dans une situation où ressources cognitives sont disponibles, où l'individu est motivé à avoir un jugement sans biais, mais que le biais perçu est supérieur au biais réel, alors l'individu auto-corrigerait son évaluation mais de manière excessive, ce qui aboutirait à une évaluation en opposition trop grande par rapport ce qu'elle aurait été en l'absence de biais. Ainsi, une manière d'expliquer pourquoi certains participants évaluaient des individus comme étant *plus* attractifs après avoir été exposés à des actrices attirantes serait que ces participants ont surestimé l'influence de l'effet de contraste sur leurs attitudes (Wegener & Petty, 1995, étude 3).

2.3.2.3 Effet de sur-correction à la suite de la correction d'informations fausses

L'effet de sur-correction apparaît assez bien établi dans plusieurs domaines de la littérature tels que la persuasion ou encore la formation d'impression (pour une revue, voir Wegener & Petty, 1997). Dans le domaine de la correction de FI, de rares études suggèrent qu'un tel effet est également possible. Le raisonnement théorique est que si un individu identifie qu'une information qui influençait son jugement ne doit plus être prise en compte, il pourrait chercher à s'auto-corriger. Toutefois, s'il surestime l'influence de la FI, sa compensation pourrait mener à un jugement excessivement en opposition à la FI. L'illustration d'un effet de sur-correction à la suite de la correction d'une FI est présentée dans la Figure 3.

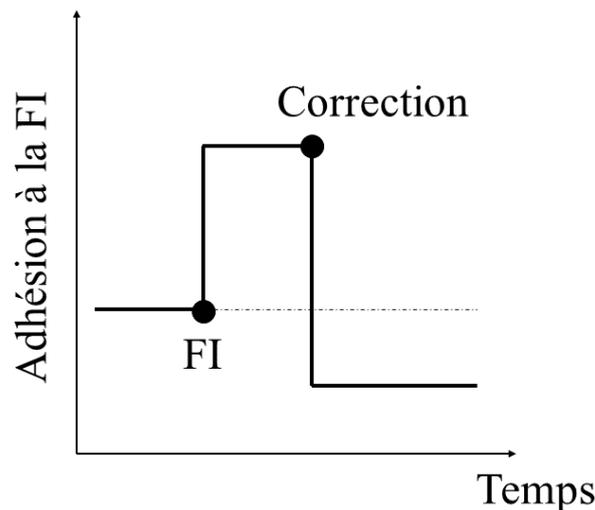


Figure 3. Illustration d'un effet de sur-correction à la suite de la correction d'une FI

Une première expérience qui supporte que la correction d'une information peut aboutir à un effet de sur-correction est celle de Sommers et Kassin (2001), bien que les auteurs n'utilisaient pas strictement une FI. Des faux jurés étaient confrontés à un enregistrement audio qui mettait en cause le suspect principal d'une affaire. Ils recevaient ensuite la consigne de ne plus prendre en compte ce document car sa mauvaise qualité le rendait finalement inutilisable. Supposant un effet de sur-correction, les participants qui avaient une haute tendance à s'engager dans des activités cognitives (mais pas ceux pour qui cette disposition était faible) avaient alors

moins tendance à trouver le suspect coupable par rapport à d'autres participants n'ayant jamais entendu parler de l'enregistrement. De même, dans une autre expérience (Cobb et al., 2013), des participants ont été exposés à une information *positive* sur un politicien qui était ensuite rétractée. Le jugement de l'homme politique par ces participants devenait alors *plus négatif* par rapport à d'autres qui n'avaient pas reçu la FI. Enfin, dans deux expériences de van Huijstee et al., (2021), un groupe de participants a reçu une information *positive* sur un hôpital qui était ensuite rétractée. Les résultats montraient que l'attitude envers l'hôpital devenait *plus négative* à la suite de ces messages. En accord avec la littérature autour de l'effet de sur-correction (Wegener & Petty, 1997), une explication possible à ces résultats est qu'en reconnaissant la possibilité qu'ils soient influencés par la FI, les participants ont excessivement compensé leurs cognitions dans le sens inverse de la FI. A notre connaissance, ces trois évidences sont toutefois les seules à avoir démontré un effet de sur-correction à la suite de la correction d'une FI. Ce phénomène appliqué à la correction des FI reste donc encore peu validé.

2.3.3 Retour ou maintien de la ligne de base

Outre l'EIC, l'effet boomerang et l'effet de sur-correction, un dernier résultat possible à la suite de la correction d'une FI est d'observer des cognitions similaires entre avant la réception de la FI et après la réception de la correction. Ce scénario est envisageable à travers deux situations qui sont le maintien de la ligne de base des cognitions ou bien un retour à la ligne de base.

2.3.3.1 Maintien de la ligne de base

Un premier moyen par lequel les cognitions après la correction peuvent se trouver similaires à celles avant la FI est simplement si la FI n'a pas eu d'influence. Dans cette situation, les cognitions resteraient les mêmes malgré la réception de la FI. La correction n'aurait donc pas d'influence non plus puisqu'elle n'aurait pas d'adhésion sur laquelle agir. Ce scénario est illustré dans la Figure 4.

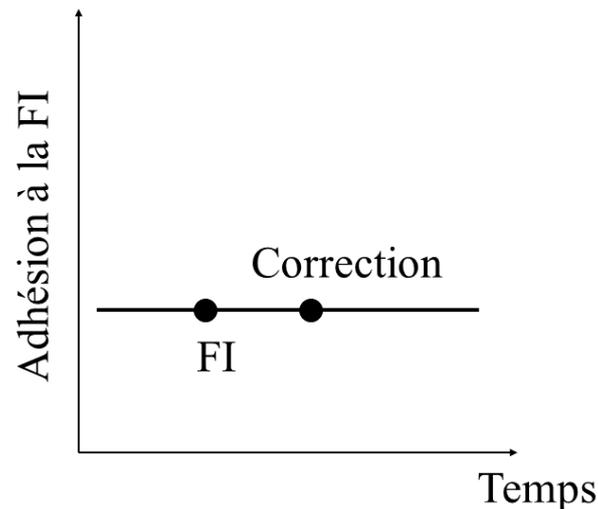


Figure 4. Illustration d'une absence d'influence de la FI

A notre connaissance, il n'existe pas de travaux qui auraient explicitement montré une absence d'influence de la FI dans le cadre de l'EIC. Il est toutefois possible d'envisager plusieurs situations propices à une absence d'influence de la FI. Une première situation qui pourrait entraîner l'absence d'influence d'une FI serait si elle n'a pas été encodée. Par exemple, un individu pourrait être exposé à une information fautive mais manquer de ressources cognitives pour l'encoder, ou être dans un environnement trop bruyant pour l'entendre, trop chargé pour la voir ou dans un langage trop différent du sien pour la comprendre. Dans ces situations, l'individu ne serait alors pas influencé par la FI bien qu'il y ait été factuellement exposé.

Une seconde situation qui pourrait expliquer l'absence d'influence d'une FI serait si elle n'est pas intégrée dans le modèle mental malgré son encodage. Comme le rappelle Rapp (2016), le fait qu'une information soit intégrée dépend de plusieurs variables. En ce sens, une expérience (Rapp et al., 2014) a exposé des participants à des textes qui contenaient ou ne contenaient pas de FI. Un texte avec FI impliquait par exemple des marins navigant grâce aux étoiles en utilisant une boussole (l'outil utilisé pour naviguer grâce aux étoiles est un sextant). Les résultats montraient ensuite que les participants commettaient plus d'erreurs à des questions

de connaissance générale (e.g., « *Quel instrument les marins utilisent-ils pour naviguer avec les étoiles ?* ») quand ils avaient été exposés à l'information fautive. Toutefois, ce taux d'erreur était moins haut lorsque le récit avait été contextualisé dans un monde de fantaisie (e.g., l'histoire se déroulait sur la lune de Jupiter) par rapport à s'il prenait place dans le monde réel, ce qui suggère que le caractère fictif du contexte dans lequel étaient reçues les informations avait modéré leur intégration. Par ailleurs, une autre caractéristique qui pourrait impacter l'intégration d'une information fautive est sa crédibilité ou sa plausibilité. Par exemple, Rapp (2016) illustre qu'une information telle que « *Saint-Petersbourg est la capitale de la Russie.* » pourrait modifier le modèle mental des individus relatif à la géographie de la Russie car elle est plausible, à l'inverse de « *Brasilia est la capitale de la Russie.* » Leur prédiction a été évaluée au sein d'une expérience dans laquelle des participants ont lu un texte qui pouvait contenir une FI plus ou moins plausible. Une histoire évoquait par exemple le voyage des premiers pèlerins anglais aux Etats-Unis sur leur navire. Un groupe lisait que le nom du bateau était le *Godspeed* (FI plausible) et un autre qu'il était le *Titanic* (FI improbable) (le vrai nom du bateau était le *Mayflower*). Des questions de connaissance générale étaient ensuite posées (e.g., « *Quel est le nom du bateau qui a transporté les pèlerins aux Etats-Unis en 1620 ?* »). Les participants répondaient moins en utilisant la FI lorsque celle-ci était improbable par rapport à quand elle était plausible, ce qui montre que la crédibilité de la FI modère aussi son intégration.

En somme, dans une situation où l'adhésion à la FI est similaire entre avant sa réception et après sa correction, une première explication possible est une absence d'influence de la FI. Cette absence d'influence est possible si la FI n'a pas été encodée, ou simplement si elle n'a pas été intégrée faute de crédibilité ou de réalisme.

2.3.3.2 Retour à la ligne de base

En plus de l'absence d'influence de la FI, un second moyen par lequel l'adhésion à une FI après sa correction pourrait être égale à son adhésion avant sa réception est si la

correction est suffisamment efficace pour revenir à une absence d'influence. Cette situation est représentée dans la Figure 5.

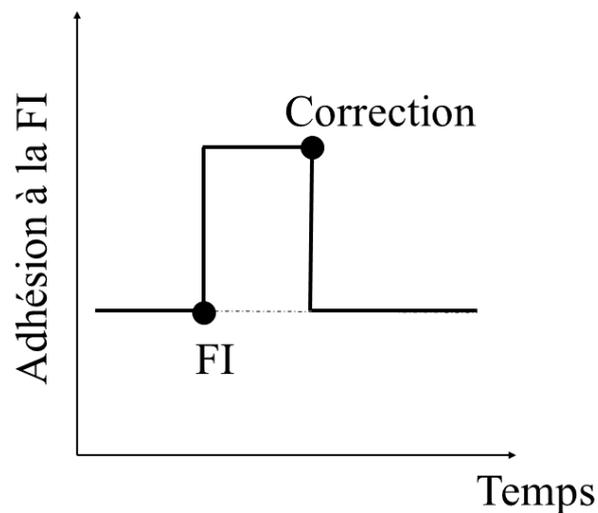


Figure 5. Illustration d'une correction efficace de la FI

Certaines recherches ont effectivement observé que dans certaines conditions, un retour à l'état de croyance initial semble possible grâce à la correction. Par exemple, dans une version du paradigme de Johnson et Seifert (1994, étude 3A), en plus d'apprendre que l'information selon laquelle l'incendie était dû à de la négligence était fausse, certains participants ont appris que la cause la cause de l'incendie était en fait criminelle. Pour les participants qui ne recevaient pas cette explication alternative, le nombre de références à la FI était supérieur par rapport à un groupe n'ayant pas reçu la FI. Toutefois, lorsqu'une explication alternative était donnée à la correction, le nombre d'inférences relatives à la FI était similaire à celui d'un groupe contrôle n'ayant pas reçu de FI. Ensemble, ces observations suggèrent que la FI avait eu une influence sur les cognitions, mais que la correction fournissant une explication alternative avait été suffisamment efficace pour évincer son influence. Une autre expérience (De keersmaecker & Roets, 2017) qui montre la possibilité d'un retour à la ligne de base a exposé des participants à des informations négatives sur un individu. Les évaluations de

l'individu étaient alors plus négatives par rapport à d'autres participants n'ayant pas reçu les informations. Dans une seconde partie, les participants apprenaient que les informations étaient fausses. Les analyses montraient alors que pour une partie des participants, l'évaluation du personnage était aussi positive qu'un groupe contrôle n'ayant pas reçu les FI (voir aussi Guillory & Geraci, 2016).

En résumé, plusieurs issues sont possibles suite à la correction d'une FI. Par effet boomerang, la correction d'une FI pourrait augmenter son adhésion. La prévalence de cet effet est toutefois remise en question au sein de la littérature. A l'inverse par effet de sur-correction, la correction pourrait réduire l'adhésion mais de manière excessive, aboutissant à des cognitions en opposition excessive par rapport à ce qui est rationnellement attendu. Cet effet reste néanmoins peu étayé au sein de la littérature relative aux FI. D'autres possibilités sont le retour ou le maintien d'une ligne de base, qui se traduisent par une absence d'influence de la FI à la suite de la correction. La riche littérature autour de l'EIC (voir le Chapitre 3) amène toutefois à penser qu'une correction pleinement efficace serait rare, et que la règle serait plutôt qu'une correction est insuffisante ou inefficace.

Chapitre 3. Réplicabilité et validité de l'effet d'influence continue

La littérature sur l'EIC montre que l'influence d'une FI peut persister malgré sa correction. A une époque où les FI constituent une problématique majeure et qu'une stratégie fréquemment employée pour lutter contre leur influence consiste à les corriger, cet effet a une importance contemporaine particulière car il remet en question l'efficacité et la pertinence de l'approche corrective. Toutefois, l'intensité de cette remise en question dépend de la validité de l'EIC. Si l'EIC est un effet valide et fréquemment observé à la suite de la correction d'une FI, alors la pertinence et l'efficacité de l'approche corrective seraient fortement remises en cause. A l'inverse, si la littérature sur l'EIC reste relativement rare et/ou montre une grande variabilité dans l'occurrence de cet effet, alors l'approche corrective resterait une stratégie pertinente pour lutter contre l'influence des FI. C'est pourquoi il apparaît nécessaire de déterminer à quel point cet effet est valide, ce qui sera l'objectif de ce chapitre.

Afin d'examiner la validité d'un effet, une première pratique consiste à étudier sa répliquabilité conceptuelle. Les répliquabilités conceptuelles désignent la répétition du test d'une hypothèse ou d'un résultat d'un précédent travail mais avec différentes méthodes (Romero, 2019). Elles sont essentielles pour les sciences car elles offrent des évidences à la validité d'un effet tout en permettant d'en évaluer le caractère généralisable (Romero, 2019). Par exemple, l'observation d'un EIC à la suite de la correction d'une FI relative à un médicament (Hamby et al., 2020) ou un politicien (Cobb et al., 2013) permet d'étayer la validité de cet effet tout en suggérant qu'il impacte les sphères sanitaires et politiques. C'est pourquoi une première partie de ce chapitre examinera l'état de la répliquabilité conceptuelle de l'EIC au sein de la littérature. Une seconde le complétera par une étude portant sur la répliquabilité conceptuelle de l'EIC avec des textes militaires.

Une autre pratique pour examiner la validité d'un effet est d'étudier sa répliquabilité proche. Les répliquabilités proches (ou directes) consistent à répéter une procédure expérimentale afin d'observer si les résultats originaux se reproduisent. Considérant qu'il est impossible de

strictement répéter une procédure (e.g., le lieu, le moment, les participants changent inévitablement) (Romero, 2019 ; Simons, 2014), une manière d'opérationnaliser ces répétitions consiste à reproduire la recherche initiale de la façon la plus proche possible (Brandt et al., 2014). L'objectif d'une troisième partie de ce chapitre sera d'étudier la réplicabilité proche de l'EIC en reprenant le paradigme de Johnson et Seifert (1994) et en observant si l'EIC se reproduit.

3.1 Réplicabilité conceptuelle de l'effet d'influence continue

Depuis l'article de Johnson et Seifert (1994), de nombreux travaux ont supporté la réplicabilité conceptuelle de l'EIC. Cette réplicabilité conceptuelle a été démontrée à travers des variations dans les procédures, dans les mesures ainsi que dans le type d'informations utilisées.

3.1.1 Variations dans les procédures

Au-delà du texte classique de l'incendie d'un entrepôt (Johnson & Seifert, 1994), de nombreux autres textes écrits employant une série de messages ont montré un EIC. Johnson et Seifert (1994, étude 3B) ont eux-mêmes démontré cet effet avec un autre texte dans lequel un fils de famille était initialement accusé d'être à l'origine du cambriolage de sa maison avant d'apprendre qu'il n'était plus suspecté. Dans d'autres expériences, des participants ont lu qu'un accident de minibus impliquait des personnes âgées (Ecker et al., 2010 ; Kan et al., 2021 ; O'Rear & Radvansky, 2020) ou qu'une compagnie pharmaceutique avait causé la mort de poissons en rejetant des déchets dans une rivière (Sanderson et al., 2021), avant d'apprendre que ces informations étaient fausses. Dans tous ces travaux et dans de nombreux autres (Autry & Duarte, 2021 ; Connor Desai et al., 2020 ; Ithisuphalap et al., 2020 ; Kan et al., 2021 ; Rich et al., 2022), les résultats montraient que l'influence de la FI persistait malgré sa correction.

Deuxièmement, au-delà de l'exposition à des textes constitués d'une série de messages, l'EIC a été observé avec des textes de formats différents. Par exemple, il a été observé

avec des informations présentées sous la forme de titres d'articles de journaux (Drummond et al., 2020) ou bien sous la forme de textes complets (McIlhiney, Gignac, Ecker, et al., 2022 ; McIlhiney, Gignac, Weinborn, et al., 2022 ; Nyhan et al., 2013 ; O'Rear & Radvansky, 2020 ; Rich et al., 2022 ; Sanderson et al., 2021 ; Wood & Porter, 2016). Il a également été observé lorsque les messages étaient présentés sous forme de tweets (Butler et al., 2023 ; Ecker, Lewandowsky, et al., 2020 ; Ecker, O'Reilly, et al., 2020 ; Kan et al., 2021 ; Vraga & Bode, 2020) ou encore de publications sur Facebook (Brashier et al., 2021).

Troisièmement, plutôt que des informations sur papier, l'EIC s'observe aussi lorsque les informations sont présentées sous la forme d'un jeu en réalité virtuelle (Sanderson et al., 2022), sur un ordinateur (Butler et al., 2023 ; Connor Desai & Reimers, 2019 ; Cook & Lewandowsky, 2016 ; Jin et al., 2022) ou encore un téléphone portable (Xu et al., 2020). Il est aussi relevé avec des informations sous format audio (Gordon et al., 2017 ; Sanderson et al., 2023). Sanderson et al. (2022) ont par exemple montré un EIC en invitant leurs participants à écouter une radio qui relatait plusieurs faits divers comportant des FI subséquemment corrigées.

Enfin, de nombreuses autres modalités du paradigme classique ont été manipulées dans la littérature, telles que la latence entre la FI et la correction (e.g., Thorson 2016) ou entre la lecture des informations et la complétion des mesures (e.g., Ecker et al., 2011), la crédibilité du désinformateur (e.g., Butler et al., 2022) ou du correcteur (e.g., Guillory & Geraci, 2013) ou encore la disponibilité des ressources attentionnelles au moment de l'encodage des informations (e.g., Ecker et al., 2011). Dans la plupart des cas (voir Chapitre 4), ces variations dans la méthode n'ont pas empêché l'EIC, supportant ainsi la répliquabilité conceptuelle de cet effet.

3.1.2 Variations dans les mesures

Dans le paradigme classique, la variable dépendante consistait à évaluer les inférences que produisaient les participants dans leurs réponses à des questions inférentielles ouvertes (Autry & Duarte, 2021 ; Ecker, Lewandowsky, & Apai, 2011 ; Ecker & Antonio, 2021). Dans

l'étude de Johnson et Seifert par exemple (1994), après avoir appris que l'incendie n'était finalement pas dû à des matériaux inflammables stockés de façon négligente, une question inférentielle ouverte était « *Sur quel aspect de l'incendie la police pourrait souhaiter d'enquêter ?* ». Chaque référence explicite à des thèmes de négligence ou de matériaux inflammables dans la réponse à cette question était alors comptée comme étant une référence à la FI.

Certaines expériences ont choisi de modifier les questions inférentielles ouvertes en questions inférentielles fermées (e.g., McIlhiney et al., 2022 ; Sanderson et al., 2022) ou de faire usage des deux types (Connor Desai & Reimers, 2019 ; Ithisuphalap et al., 2020). Par exemple, Connor Desai et Reimers (2019) ont repris le texte classique de Johnson et Seifert (1994) puis ont adapté des questions inférentielles ouvertes (étude 1A, étude 2A) en questions inférentielles fermées (étude 1B, étude 2B). Ainsi, à la question « *Sur quel aspect de l'incendie la police pourrait souhaiter d'enquêter ?* », quatre modalités de réponse dont une faisait référence à la FI (i.e., « *des matériaux inflammables stockés de façon négligente* ») étaient proposées. Les participants attribuaient des points de plausibilité à chacune des propositions et le total des points attribués aux modalités faisant référence à la FI était récolté. Par ailleurs, d'autres études ont choisi d'utiliser des échelles de Likert. Par exemple, dans une recherche dans laquelle ils apprenaient que la mort de poissons dans une rivière n'était finalement pas due au rejet de déchets chimiques par une compagnie pharmaceutique, des participants répondaient ensuite sur des échelles en 11 points (0 : « *complètement en désaccord* », 10 : « *complètement d'accord* ») à des affirmations telles que « *La compagnie pharmaceutique proche de la rivière devrait mener une enquête interne et revoir ses procédures.* » (Sanderson et al., 2021).

Une autre déviation réalisée par rapport aux questions inférentielles ouvertes a été d'utiliser des mesures d'adhésion à la croyance. La caractéristique des mesures d'adhésion à la croyance est d'évaluer l'élément cognitif impacté par la FI de façon directe. Elles se

différencient ainsi des questions inférentielles qui se caractérisent quant à elle par une évaluation indirecte de l'adhésion à la FI. La distinction entre ces deux types de mesure peut être illustrée avec une étude de Ithisuphalap et al. (2020). Dans cette étude, un texte indiquait aux participants qu'une femme était tombée malade à la suite d'une soirée en boîte de nuit car son verre avait été empoisonné, avant d'informer que sa maladie était finalement due à une bactérie sans lien avec sa soirée. Les questions inférentielles étaient par exemple d'évaluer l'accord avec la proposition selon laquelle la boîte de nuit devrait instaurer une politique où seules les consommations à la bouteille seraient autorisées. La mesure d'adhésion à la croyance consistait quant à elle à mesurer l'intensité avec laquelle les participants pensaient que la femme était tombée malade parce que son verre avait été empoisonné.

L'importance de distinguer les questions inférentielles et la mesure d'adhésion à la croyance est confirmée par deux types de résultats. Premièrement, plusieurs recherches ont observé que la corrélation entre ces deux variables n'était pas parfaite. Par exemple, dans l'étude de Ithisuphalap et al. (2020) évaluant la persistance de l'information selon laquelle la jeune femme aurait été empoisonnée, et à l'instar d'autres travaux (Rich et al., 2022 ; Rich & Zaragoza, 2015 ; Susmann & Wegener, 2021), seule une corrélation modérée était relevée entre la mesure d'adhésion à la croyance et le score aux questions inférentielles (voir cependant Swire et al., 2017 pour une corrélation forte). Deuxièmement, ces deux types de mesure suscitent parfois des résultats différents. Par exemple, Butler et al. (2022) ont observé que le nombre de *likes* associés à une FI sur Twitter augmentait l'EIC lorsqu'il était mesuré par l'adhésion à la croyance mais pas quand il l'était par des questions inférentielles. De même, Ecker et al. (2020) ont relevé un effet boomerang dû à la familiarité (non-répliqué par la suite) en employant des questions inférentielles, mais pas avec une mesure d'adhésion à la croyance.

Enfin, au-delà des questions inférentielles et de la mesure d'adhésion à la croyance, de rares recherches ont montré un EIC par le biais de mesures comportementales ou d'intention

comportementale. Par exemple, dans une expérience (MacFarlane et al., 2018), des participants ont été informés de l'inefficacité des compléments multivitaminés sur la santé tandis que d'autres n'avaient reçu aucune correction. Un EIC était montré en comparant quel montant les participants étaient prêts à miser dans une vente aux enchères fictive de ce produit. De même, dans une autre étude (Hamby et al., 2020), un groupe de participants a été informé dans un texte que le kombucha avait la capacité d'améliorer les performances cognitives, avant d'apprendre finalement que ce qu'ils avaient lu était faux. Un EIC était évalué en offrant aux participants, à la fin de l'expérience, la possibilité de choisir entre plusieurs boissons dont le kombucha. Ces deux travaux (Hamby et al., 2020 ; MacFarlane et al., 2018) sont à notre connaissance les seuls à avoir démontré un EIC avec des mesures de type comportemental. Le constat d'un manque d'études usant de variables de type comportemental dans le cadre de l'EIC s'accorde avec la littérature plus générale autour des FI, au sein de laquelle plusieurs chercheurs avaient déjà noté que l'utilisation de telles mesures était rare (Courchesne et al., 2021 ; de Saint Laurent et al., 2022 ; Murphy et al., 2023).

3.1.3 Variations dans les domaines d'application

En plus de variations dans les procédures et les mesures utilisés, beaucoup de travaux ont supporté la répliquabilité conceptuelle de l'EIC en relevant cet effet dans de nombreuses sphères. Par exemple dans le domaine politique, la persistance de FI malgré leur correction peut être illustrée avec le mouvement des *birthers* qui contestait la nationalité américaine de Barack Obama. Malgré la diffusion publique de son acte de naissance, l'adhésion à cette FI a persisté durant son mandat (Tuchman, 2011). De façon plus expérimentale, une expérience (Cobb et al., 2013) a indiqué à des participants qu'un politicien était opposé à des mesures favorables à la population puis que la source d'information avait en fait confondu ce politicien avec un autre. Malgré cela, les sujets continuaient de percevoir plus négativement ce dernier par rapport à si aucune FI n'avait été reçue. De la même manière, plusieurs études ont informé des participants

qu'un politicien aurait détourné des fonds (Ecker & Ang, 2019), été vu en train de recevoir un pot-de-vin (Guillory & Geraci, 2013, 2016) ou aurait été financé par un criminel (Thorson, 2016). Malgré la correction de ces informations, de plus nombreuses références aux thèmes de fraudes (Ecker & Ang, 2019 ; Guillory & Geraci, 2013, 2016), une attitude plus négative envers l'homme politique (Thorson, 2016) et une moindre tendance à voter pour ce dernier (Guillory & Geraci, 2013) étaient observées par rapport à d'autres participants n'ayant pas reçu de FI (voir aussi Brashier et al., 2021).

Outre le domaine politique, le sujet du climat semble également sensible à l'EIC. Cook et Lewandowsky (2016) ont par exemple exposé leurs participants à un message mettant en avant le consensus d'experts au sujet de l'origine anthropogène du réchauffement climatique et ont relevé que malgré ce message, une partie notable continuait d'adhérer à des FI sur ce sujet. De même, dans une autre expérience (Drummond et al., 2020), un groupe a lu des informations fausses sur le sujet du réchauffement climatique (e.g., les résultats des études sur le réchauffement climatique seraient fabriquées). Le caractère factice de ces informations était ensuite révélé. Malgré la correction, la croyance en l'existence du réchauffement climatique restait plus basse par rapport à un groupe contrôle n'ayant pas reçu les FI (voir aussi Swire-Thompson et al., 2021).

Certaines recherches permettent aussi d'avancer que l'EIC est une problématique pour le domaine de la santé, bien qu'elles n'aient pas strictement contrôlé l'exposition à une FI. Une étude a par exemple interrogé des individus qui avaient indiqué avoir partagé au moins une FI relative au COVID-19 sur les réseaux sociaux. Leur adhésion à des FI sur ce sujet était évaluée et il leur était également demandé s'ils avaient été informés par un tiers qu'ils avaient partagé une FI. Les résultats ont montré que ceux qui déclaraient avoir été corrigés montraient autant de fausses croyances autour du COVID-19 par rapport à ceux qui ne l'avaient pas été (Bode & Vraga, 2021). De la même manière, une autre étude a exposé des participants à un message

démontrant que les compléments multivitaminés n'avaient aucun effet bénéfique sur la santé. Ils étaient ensuite invités à miser une somme d'argent proportionnelle à la valeur qu'ils accordaient au produit. Supportant un manque d'efficacité de la correction, la somme que ces participants étaient prêts à payer pour ce produit était similaire par rapport à d'autres n'ayant pas reçu de correction (MacFarlane et al., 2018).

Par ailleurs, l'EIC est aussi une problématique importante pour le domaine de la justice (Black, 2018). De premières illustrations sont portées par des expériences déjà mentionnées relatives au domaine politique, qui montraient qu'une information indiquant qu'un politicien aurait détourné des fonds (Ecker & Ang, 2019) ou reçu un pot-de-vin (Guillory & Geraci, 2013) peut persister malgré sa correction. Également, d'autres recherches étayaient la validité de l'EIC dans la sphère judiciaire, bien qu'elles n'aient pas utilisé d'informations strictement fausses. Dans ces expériences, des participants prenaient un rôle fictif de jurés et devaient juger de la culpabilité d'un accusé. Certains apprenaient que l'accusé avait réalisé des aveux (Kassin & Sukel, 1997), ou bien que des articles de presse (Fein et al., 1997, étude 1), qu'un témoignage (Fein et al., 1997, étude 2) ou encore qu'un enregistrement audio l'incriminaient (Sommers & Kassin, 2001). Les participants recevaient ensuite la consigne de ne plus prendre en compte ces éléments, respectivement car les aveux avaient été réalisés sous pression, que les articles de presse ne sont pas des preuves, que le témoignage était un ouï-dire et que l'enregistrement avait été obtenu de façon illégale. Malgré cette consigne, leur tendance à trouver le suspect coupable demeurait supérieure par rapport à s'ils n'avaient pas reçu ces informations. Enfin, plus directement en lien avec l'EIC et le domaine de la justice, une expérience a décrit à des participants un personnage qui aurait été arrêté pour vente de drogue, avant de leur apprendre que cette dernière information était finalement fausse. Malgré la correction, une partie de ces participants continuait de garder une perception plus négative du personnage par rapport à d'autres n'ayant pas reçu la FI (de Keersmaecker & Roets, 2017).

En outre, dans le domaine économique, plusieurs recherches ont montré un EIC avec des publicités mensongères. Par exemple, une expérience a observé que des participants exposés à une publicité positive concernant un supermarché continuaient d'avoir une perception plus positive de cette enseigne même lorsque le caractère factice de la publicité leur était révélé (Misra, 1992, étude 3). De même, dans l'étude de Hamby et al. (2020), des participants qui apprenaient que le kombucha n'avait finalement pas la capacité d'améliorer les performances cognitives continuaient malgré la correction de préférer ce produit par rapport à s'ils n'avaient pas reçu la FI.

Enfin, un autre domaine pour lequel la validité de l'EIC a été abordée, bien que de manière indirecte, est le domaine militaire. Dans une recherche, des participants ont initialement appris qu'un crash aérien aurait été causé par une attaque terroriste, avant d'apprendre qu'il résultait d'une défaillance technique. Malgré la correction, ils restaient plus enclins à attribuer l'évènement à une attaque terroriste par rapport à des participants n'ayant pas reçu la FI (Ecker, Lewandowsky, & Apai, 2011). Également, dans une autre étude, des analyses descriptives ont relevé que des américains exposés à des FI sur la guerre en Irak avaient tendance à y adhérer même s'ils savaient qu'elles avaient été rétractées (Lewandowsky et al., 2005). Enfin, à travers une méthodologie qualitative, Prasad et al. (2009) semblaient observer que la plupart des individus qui avaient exprimé une croyance erronée concernant l'implication de Saddam Hussein dans les attaques du 11 septembre ne manifestaient pas de modification de leurs cognitions lorsqu'ils étaient confrontés à des éléments qui la contredisait. Ces recherches suggèrent qu'un EIC pourrait se produire dans le monde des armées. Elles n'offrent toutefois pas une réponse directe à cette question car les textes utilisés n'étaient pas pleinement militaires (Ecker, Lewandowsky, & Apai, 2011) ou bien ne suivaient pas une méthodologie expérimentale (Lewandowsky et al., 2005 ; Prasad et al., 2009). De ce fait, la répliquabilité conceptuelle de l'EIC dans le domaine militaire reste limitée.

3.2 Expérience 1. Réplicabilité conceptuelle au domaine militaire

Dans le contexte d'un manque de répliquabilité conceptuelle de l'EIC dans le domaine militaire, l'objectif de l'Expérience 1 a été d'examiner l'occurrence de cet effet en utilisant des textes militaires et une approche expérimentale. Des participants tout-venant ont été exposés à quatre textes décrivant des histoires fictives vécues par un officier militaire confronté à une décision. Deux des textes comportaient une FI qui était ensuite corrigée (version FI) et les deux autres n'en contenaient pas (version contrôle). Après chaque texte, les participants avaient pour consigne de répondre à des questions inférentielles. Une variable de type comportemental consistant à évaluer la décision que les participants prendraient s'ils étaient à la place de l'officier militaire a aussi été utilisée. Les réponses étaient comparées entre les deux versions de texte de façon intraindividuelle. En accord avec l'EIC, nous avons émis l'hypothèse que les textes en version FI devraient susciter plus de décisions et d'inférences en lien avec la FI par rapport aux textes en version contrôle. A titre exploratoire, nous avons aussi examiné si la confiance des participants dans leur choix ainsi que leur sentiment d'importance et de difficulté dans sa réalisation variait en fonction des textes.

3.2.1 Méthode

Le projet de recherche a été validé par le Comité d'éthique en sciences comportementales de l'Université de Lille (référence 2022-582-S103) et a été réalisé en accord avec les standards éthiques de la Déclaration d'Helsinki de 1964. Le consentement éclairé a été obtenu pour tous les participants inclus dans cette étude. Le matériel utilisé est accessible en ligne avec le lien suivant <https://osf.io/bcfej/>. Cette expérience a donné lieu à une publication dans la revue *Applied Cognitive Psychology* (Laurent et al., 2024).

3.2.1.1 Participants

Le critère d'inclusion pour cette expérience était d'être âgé d'au moins 18 ans et de ne pas avoir de troubles diagnostiqués de lecture. En cohérence avec des études précédentes sur

l'EIC (e.g., Chang et al., 2019 ; Ecker et al., 2011, 2017 ; Johnson & Seifert, 1994), nous avons décidé de tester environ 60 participants. Un total de 62 participants a été recruté. Les données d'un participant ont été retirées car sept (77,78 %) de ses réponses aux questions de rappel étaient incorrectes et douze (100 %) de ses réponses aux questions inférentielles étaient la réponse par défaut. Les décisions d'un autre participant ont été enlevées des analyses car sa latence de décision que relevait automatiquement le logiciel supportant l'expérience était de 0 seconde, ce qui suppose l'occurrence d'un problème technique. Toutes ses autres réponses ont été incluses dans les analyses. Ainsi, l'échantillon final était de 61 et de 60 pour l'analyse des prises de décision. Parmi les 61 participants, 39 s'identifiaient de genre féminin et un s'identifiait de genre « autre ». Par ailleurs, 52 d'entre eux (83,87 %) avaient un niveau d'études supérieur au baccalauréat. L'âge moyen était de 34 ans ($\sigma = 13,999$).

3.2.1.2 Matériel

3.2.1.2.1 Textes

Quatre textes ont été créés dans le cadre de cette expérience. Afin d'assurer leur validité militaire, ces textes ont été écrits par un premier officier de la Marine nationale puis validés par un second. De plus, ils sont inspirés d'évènements réels, tels que le vol Iran Air 655 (Rochlin, 1991), les leurres utilisés par Gengis Khan (McKey, 2014 ; Rouse, s. d.) et dans l'Opération Fortitude (Anderson, 2010 ; De Witte Parra, 2019 ; Kornegay, 2013 ; Tavares, 2001) ou encore le faux message reçu par l'OTAN (NATO, 2020) (voir partie 1.3). La structure des textes était standardisée. Ils étaient écrits au pluriel de politesse (« vous ») et au présent. Ils décrivaient des histoires fictives vécues par un officier de la Marine nationale faisant face à une décision. Ils se composaient de 18 messages, titre inclus. Les messages étaient écrits en noir et apparaissaient au centre de l'écran. L'arrière-plan était une image qui variait en fonction de chaque texte. Chaque texte débutait par un écran présentant le titre, suivi par un écran indiquant au participant qu'il n'avait pas de contrainte de temps pour lire chaque message mais qu'il ne

pourrait pas retourner en arrière dans sa lecture. Le troisième message exposait la situation initiale (e.g., « *Vous êtes le commandant d'un navire de combat [...]* »). Le quatrième suggérait aux participants de porter une attention particulière sur certains éléments, qui correspondaient en fait aux informations liées à la prise de décision à venir (e.g., « *Attention, [...] votre mission consiste à surveiller l'espace aérien. Ainsi, vous vous devez de prendre avec une considération particulière toutes informations relatives à cet espace.* ») Le dixième message variait en fonction de la version du texte, laquelle pouvait être une version contrôle ou une version FI. Dans la version FI, une information importante pour la prise de décision à venir était fournie mais était ensuite rétractée dans le quinzième message (i.e., le message correctif). Dans la version contrôle, la FI était remplacée par une information neutre écrite de sorte à ne pas influencer la décision, et le quinzième message était le même que dans la version FI. Tous les autres messages étaient des détails relatifs à l'histoire. L'avant-dernier message résumait les éléments du texte et le dernier introduisait la décision à suivre (i.e., « *Vous devez prendre une décision maintenant.* ») La standardisation des textes et l'illustration de l'un d'entre eux sont présentées dans la Figure 6.

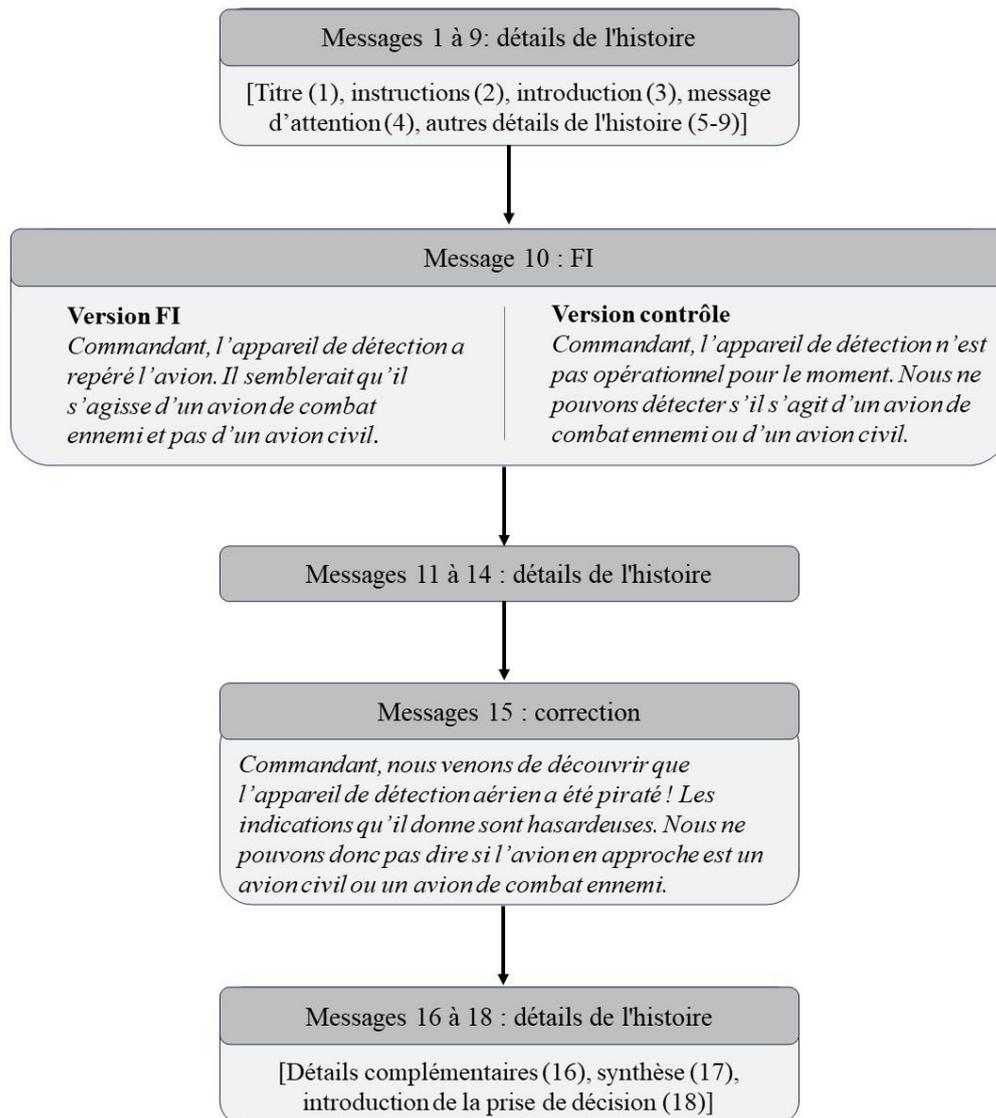


Figure 6. Illustration de la standardisation des textes

3.2.1.2.2 Mesures

Choix. Après chaque texte, les participants devaient indiquer la décision qu'ils prendraient s'ils étaient à la place de l'officier militaire. Ils choisissaient entre deux options qui leur étaient proposées (e.g., « Tirer » ou « Ne pas tirer »). Les participants indiquaient leur choix en utilisant la flèche gauche ou la flèche droite du clavier. Pour chaque texte, la décision prise avec la flèche droite indiquait un choix cohérent avec la FI (e.g., « Tirer ») et celle prise avec la flèche gauche une décision qui allait dans le sens inverse de la FI (e.g., « Ne pas tirer »).

Pour chaque participant, le nombre de décisions cohérentes avec la FI était calculé pour les textes en version FI et ceux en version contrôle (minimum = 0, maximum = 2).

Confiance, importance et difficulté dans le choix. Les participants précisaient la confiance, l'importance et la difficulté qu'ils attribuaient à la décision sur une Échelle Visuelle Analogique (EVA). L'échelle allait de 0 à 100 et sa légende variait selon la mesure (respectivement : « *Pas du tout/Tout à fait confiance* » ; « *Pas du tout/Extrêmement important* » ; « *Pas du tout/Extrêmement inconfortable* »). La réponse par défaut sur les EVA était de 50.

Questions inférentielles. Pour chaque texte, trois questions inférentielles invitaient les participants à extrapoler leur analyse de la situation en exprimant leur perception sur des éléments qui n'avaient pas été explicitement cités dans le texte (e.g., « *Considérant toute la situation, quelles sont vos chances de survie ?* »). Ils se positionnaient sur une EVA allant de 0 à 100. Les réponses étaient codées de sorte à ce que plus le score était haut, plus celui-ci supposait une référence à la FI. Par conséquent, le codage d'une réponse pour laquelle un haut score indiquait une faible référence à la FI était inversé (le nombre 100 était soustrait par le score brut).

Questions de rappel. Deux questions de rappel étaient posées après chaque texte. Une question portait sur un élément du texte (e.g., « *Quel est le nom de l'aéroport duquel a décollé l'avion ?* ») et l'autre portait sur la correction (e.g., « *Quel est le problème avec l'appareil de détection anti-aérien ?* »). Les participants choisissaient une réponse parmi quatre modalités qui leur étaient proposées. Une bonne réponse était cotée 1 et une mauvaise 0. Les résultats ont confirmé la compréhension des textes car pour chacun d'entre eux, les deux questions de rappel ont été réussies par au moins 80 % des participants. Également, le score moyen sur les huit questions de rappel, indépendamment des textes, était de 7,475 ($\sigma = 0,808$).

3.2.1.3 Procédure

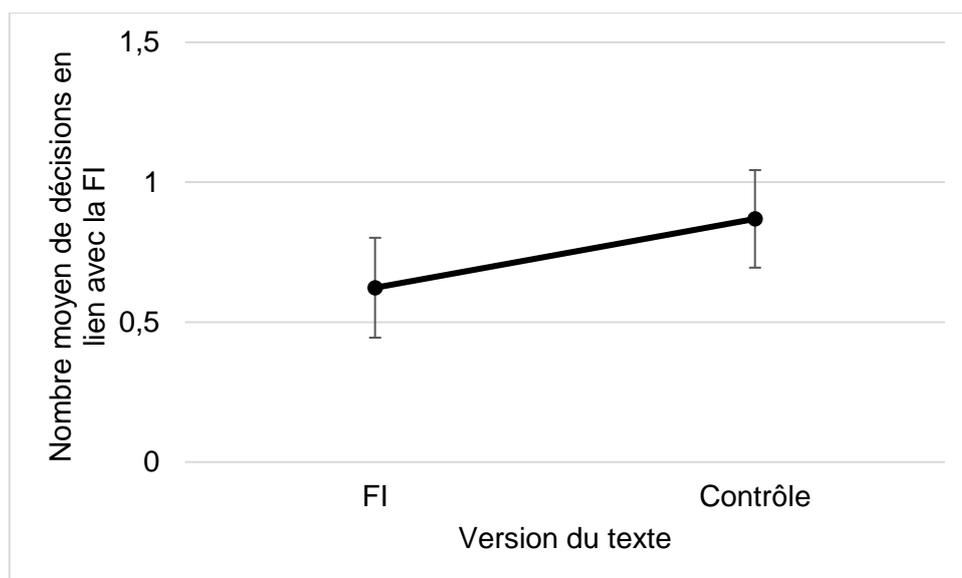
La procédure était informatisée. Elle était supportée par le logiciel Inquisit dans sa version web. Les participants étaient recrutés par le biais des réseaux sociaux ou d'affiches. Ils accédaient à l'étude avec un lien Internet. Le protocole commençait par la lecture d'une lettre d'information qui indiquait que l'expérience avait pour objectif d'étudier la prise de décision dans des scénarios militaires. Les participants étaient informés qu'ils auraient à lire quatre textes décrivant des scénarios militaires puis à répondre à des questions et imaginer prendre une décision en lien avec chaque situation. S'ils ne s'opposaient pas à prendre part à la recherche, ils signalaient leur genre, leur âge, leur niveau d'études et leur expérience militaire le cas échéant. Chaque participant était exposé aux quatre textes, dont deux dans une version FI et deux dans une version contrôle. L'ordre et la version des quatre textes étaient contrebalancés. Les participants passaient chaque message en appuyant sur la touche « espace » de leur clavier. Les variables en lien avec un texte étaient présentées immédiatement après sa lecture dans l'ordre suivant : décision, confiance, importance et difficulté dans la décision, questions inférentielles puis questions de rappel. Il n'y avait pas de limite de temps pour compléter les mesures. Une tâche intermédiaire était insérée entre chaque texte. Elle consistait à compter le nombre d'occurrence d'un chiffre au sein d'un tableau de 9*9 cases, chacune composée d'un chiffre défini de façon aléatoire. L'expérience se terminait par un texte de debriefing qui révélait le but réel de l'expérience. Il était alors demandé aux participants s'ils acceptaient que leurs données soient conservées. Ceux qui refusaient étaient retirés de l'analyse des données.

3.2.2 Résultats

Les scores moyens pour chaque variable dans les deux textes en version FI ont été comparés de manière intraindividuelle à ceux des textes en version contrôle. Afin de contrôler l'influence des versions sur la compréhension des textes, le nombre de bonnes réponses aux questions de rappel a été comparé entre les versions FI et contrôle. Un test de Wilcoxon a été

utilisé en raison du manque de normalité de la distribution des différences ($W = 0,754$, $p < 0,001$). L'analyse ne montrait pas de différence de compréhension entre les versions de texte ($W = 77$, $p = 0,704$).

Concernant le nombre de décisions relatives à la FI, les différences entre les deux versions ont été comparées avec un test de Wilcoxon en raison du non-respect de la normalité de la distribution des différences ($W = 0,902$, $p < 0,001$). Le test a montré une différence significative entre les deux versions ($W = 216$, $p = 0,05$, corrélation de rangs bisériés = $-0,351$). Plus précisément, les textes en version FI étaient associés à une *plus faible* tendance ($\bar{x} = 0,623$, $\sigma = 0,711$) à prendre des décisions cohérentes avec la FI par rapport aux textes en version contrôle ($\bar{x} = 0,869$, $\sigma = 0,695$). Ce résultat est illustré au sein de la Figure 7.



Note. Les barres d'erreur sont des intervalles de confiance, $\alpha = 0,05$.

Figure 7. Nombre moyen de décisions cohérentes avec la FI en fonction de la version des textes

Concernant les questions inférentielles, les analyses ont montré que le score moyen aux trois questions inférentielles pour les textes en version contrôle ($\bar{x} = 47,3$, $\sigma = 11,4$) n'était

pas significativement différent ($t(60) = -1,28, p = 0,206$) par rapport aux textes en version FI ($\bar{x} = 44,5, \sigma = 11$).

Enfin, les analyses ont examiné si la confiance, l'importance, et la difficulté perçues dans les décisions pour les textes en version contrôle (respectivement : $\bar{x} = 58,4, \sigma = 21,5$; $\bar{x} = 82,7, \sigma = 14,3$; $\bar{x} = 56,9, \sigma = 24,7$) étaient significativement différentes par rapport aux textes en version FI (respectivement : $\bar{x} = 58,5, \sigma = 19,4$; $\bar{x} = 83,6, \sigma = 14,5$; $\bar{x} = 55,9, \sigma = 21,5$). La distribution des différences n'était pas normale pour les mesures de l'importance ($W = 0,959, p = 0,040$) et de la confiance ($W = 0,961, p = 0,052$). Les différences entre les versions n'étaient pas significatives pour l'importance ($W = 779, p = 0,580$), la confiance ($W = 957, p = 0,589$) et la difficulté perçues ($t(60) = -0,312, p = 0,756$). Notons que les analyses descriptives pour les quatre textes, indépendamment de leur version, montraient que les scores moyens des participants pour la confiance et la difficulté perçues étaient proches du milieu de la mesure (respectivement : $\bar{x} = 58,463, \sigma = 27,595$; $\bar{x} = 56,430, \sigma = 31,157$) et que l'importance perçue était haute ($\bar{x} = 83,168, \sigma = 18,112$).

3.2.3 Discussion

L'objectif de l'Expérience 1 était d'examiner l'occurrence d'un EIC dans le domaine militaire en utilisant une approche expérimentale. Nous avons émis l'hypothèse d'un EIC et n'avons pas prédit d'effet de sur-correction. Nous avons créé quatre textes militaires pour cette expérience et les analyses descriptives ont confirmé leur bonne compréhension. Les participants lisaient ces textes dans une version FI ou contrôle et indiquaient ensuite leur décision puis répondaient à des questions inférentielles. Contrairement à notre hypothèse initiale, nous n'avons pas observé d'EIC. A la place, un effet de sur-correction a été relevé sur la prise de décision. Quand les participants lisaient un texte comportant une FI corrigée, ils étaient ensuite *moins* enclins à prendre des décisions cohérentes avec la FI par rapport à si le texte ne

comportait pas de FI. Concernant les réponses aux questions inférentielles, les résultats n'ont pas révélé de différence significative entre les deux versions de textes.

Une première observation notable est l'absence d'EIC observé sur les réponses aux questions inférentielle, qui va à l'encontre de la littérature qui suggère qu'une FI persisterait malgré sa correction (Lewandowsky et al., 2012 ; Walter & Murphy, 2018 ; Walter & Tukachinsky, 2020). Une explication à l'absence d'EIC pourrait être que les FI n'ont pas eu d'influence dans les textes en version FI. Toutefois, l'hypothèse d'un maintien de la ligne de base n'apparaît pas cohérente avec l'effet de sur-correction observé qui sera ensuite discuté, mais qui suggère déjà que la FI aurait eu une influence puisque les participants auraient été à son encontre. Par conséquent, une hypothèse plausible est que les FI ont eu une influence, mais que la correction a été suffisamment efficace pour permettre au minimum un retour des cognitions à la ligne de base.

Une explication possible au retour des cognitions à la ligne de base réside dans plusieurs éléments de la méthodologie. L'importance perçue de la tâche (Chaiken & Maheswaran, 1994), l'utilisation de la seconde personne (Burnkrant & Unnava, 1989), la conjugaison des verbes au présent (Tal-Or & Cohen, 2010), l'absence de limite de temps pour compléter les mesures (Bago et al., 2020), le fait d'imaginer prendre un rôle avec des responsabilités (Ginossar & Trope, 1987 ; Zukier & Pepitone, 1984) ou encore l'incitation à porter attention aux informations (Rapp & Kendeou, 2007) sont des éléments employés dans cette étude que la littérature identifie comme augmentant la motivation des individus à analyser les informations. Or, il est possible qu'une analyse accrue des informations contribue à réduire l'EIC. En effet, une analyse accrue des informations pourrait faciliter la détection du conflit entre la FI et la correction et donc réduire le risque que l'EIC survienne en raison d'un défaut d'activation de la FI au moment de la correction. De plus, en mobilisant des ressources cognitives, elle pourrait aussi faciliter l'intégration de la correction. Enfin, la motivation à

analyser les informations pourrait se poursuivre au moment de répondre aux mesures, ce qui pourrait faciliter la récupération de la correction (voir la partie 4.5 pour une étude approfondie de l'effet de l'analyse des informations sur l'EIC). Ainsi, il est possible que les nombreux éléments de la méthodologie contribuant à une analyse accrue des informations aient mené à la correction efficace observée dans cette recherche.

En outre, il doit être noté que plusieurs éléments limitent le caractère militaire de cette expérience. Une limite est par exemple que la plupart des participants n'avaient pas d'expérience militaire. Le choix de recruter des participants tout-venant est justifiable par le fait que dans un contexte de guerre réel, ce sont justement les civils d'aujourd'hui qui seraient appelés à rejoindre l'armée et à participer aux décisions militaires. De plus, plusieurs études ont observé que l'EIC n'était pas sensible à plusieurs variables démographiques telles que l'âge, le genre et le niveau d'études (Jin et al., 2022 ; Walter & Murphy, 2018). Malgré ces arguments qui réduisent l'importance de cette limite, le manque de participants militaires reste un point notable qui réduit le cadre conceptuel militaire de ce travail. Une autre limite au caractère militaire de cette étude est qu'elle manipulait des décisions et des scénarios fictifs plutôt que des situations réelles. Or, les comportements réels peuvent diverger par rapport aux conduites anticipées. Par exemple, une méta-analyse (Kormos & Gifford, 2014) n'a trouvé qu'une corrélation modérée ($r = 0,46$) entre les comportements pro-environnementaux auto-rapportés et les comportements effectifs. Bien que manipuler une information et sa correction dans de vraies situations militaires poserait des problématiques éthiques et d'opérationnalisation, il reste envisageable de développer une méthode plus écologique que la lecture de textes et l'utilisation de décisions pleinement fictives. Par exemple, une expérience pourrait explorer l'effet de la correction d'une FI militaire sur la prise de décision en utilisant des *wargames* (Bourguilleau, 2023) ou encore lors d'exercices de commandement. Enfin, il doit être noté que l'absence de contrainte de temps dans la prise de décision ou encore la pleine disponibilité des ressources

attentionnelles sont d'autres éléments méthodologiques qui réduisent la validité écologique du paradigme tout en réduisant la probabilité de voir apparaître un EIC (Bago et al., 2020 ; Ecker, Lewandowsky, Swire, et al., 2011).

En raison de ces limites au caractère écologique et militaire de cette expérience, nous ne pensons pas que les résultats observés doivent étayer l'idée que dans un contexte militaire, un EIC ne serait jamais observable. Afin de réexaminer la répliquabilité conceptuelle de l'EIC dans le contexte militaire, de futures expériences pourraient bénéficier de l'emploi de méthodologies suscitant un traitement de l'information et des prises de décision qui seraient davantage semblables à des conditions réelles tout en recrutant des participants militaires.

Outre l'absence d'EIC sur les questions inférentielles, un autre résultat notable est l'occurrence d'un effet de sur-correction sur la prise de décision. Les textes comportant une FI qui était ensuite corrigée suscitaient *moins* de décisions cohérentes avec la FI par rapport aux textes qui ne comportaient pas de FI. Ce constat supporte la validité de l'effet de sur-correction à la suite de la correction d'une FI. Le support de la validité de cet effet constitue un apport théorique important dans la mesure où il est encore peu exploré au sein de la littérature (voir cependant Cobb et al., 2013 ; Sommers & Kassin, 2001 ; van Huijstee et al., 2021). Le support de la validité de cet effet a aussi des implications pratiques notables. L'effet de sur-correction montre que les cognitions des participants continuent d'être influencées par une information qui n'aurait pas dû être reçue et qui, en raison de son caractère factice, ne devrait plus être prise en compte. Par conséquent, il est probable que cet effet augmente le risque de réaliser des décisions sous-optimales et de produire des erreurs. En raison des conséquences potentielles de cet effet, de futures études sur l'effet de sur-correction à la suite de la correction d'une FI sont nécessaires pour mieux comprendre ce phénomène.

Une première perspective que pourrait explorer cette littérature serait de déterminer dans quelles circonstances un effet de sur-correction, un EIC ou bien une correction efficace se

réalisent. Dans le cadre de cette expérience, une explication possible à l'effet de sur-correction est que les variables précédemment rapportées motivant l'analyse des informations ont aussi facilité son occurrence. En effet, la motivation à analyser les informations a pu s'accompagner d'une motivation à avoir un jugement sans biais, amenant les participants à détecter l'influence potentielle de la FI, les poussant à s'auto-corriger contre cette-dernière, ce qui a pu résulter en une compensation excessive si son influence était surestimée (voir Wegener & Petty, 1997).

Par ailleurs, il peut être relevé que l'effet de sur-correction a été observé sur les prises de décision mais pas sur les réponses aux questions inférentielles. Une raison possible à cette différence est que les décisions étaient plus explicitement liées à la FI corrigée par rapport aux questions inférentielles. Par exemple, le choix de « Tirer » ou de « Ne pas tirer » était directement en lien avec la FI indiquant qu'un avion en approche était ennemi. Avant de donner leur décision, les participants ont alors pu facilement détecter que la FI pourrait influencer leur réponse, ce qui a pu les inciter à s'auto-corriger. A l'inverse, la réponse à des questions telles que « *Considérant la situation, quelles sont vos chances de survie ?* » inclue non seulement la FI mais aussi d'autres éléments du texte tels que la dangerosité de la zone de navigation. Le lien entre les questions inférentielles et la FI étant moins saillant, cette mesure est probablement moins propice à ce que les participants détectent que la FI corrigée pourrait impacter leur réponse, ce qui réduirait la probabilité d'une auto-correction.

Enfin, il y a deux limites notables relatives à l'observation de l'effet de sur-correction dans le présent travail. D'une part, cet effet n'était pas attendu au sein de nos hypothèses, ce qui augmente le risque de faux positifs et appelle donc à le prendre en compte avec prudence. D'autre part, la littérature sur l'effet de sur-correction dans le contexte de la correction des FI reste rare. En plus d'appeler à nouveau à prendre ce résultat avec précaution, cette seconde limite ajoute que l'interprétation théorique qui en est faite gagnerait à être encore étayée scientifiquement.

En somme, cette expérience a permis de mettre en avant deux résultats importants. Premièrement, elle rappelle que les corrections pleinement efficaces semblent possibles, comme le suggère l'absence de différence observée dans les réponses aux questions inférentielles entre les versions de texte. Deuxièmement, elle supporte la validité de l'effet de sur-correction en montrant que des textes militaires comportant une FI corrigée peuvent susciter davantage de décisions en opposition à la FI par rapport à s'ils ne comportent pas de FI corrigée. En raison de limites de cette expérience et de la littérature, de futurs travaux apparaissent nécessaires pour mieux comprendre l'effet de sur-correction et pour réévaluer la répliquabilité conceptuelle de l'EIC au domaine militaire.

3.3 Expérience 2. Répliquabilité proche avec une population francophone

Dans le cadre de l'Expérience 1, l'absence d'EIC a été expliquée par une analyse accrue des informations par les participants plutôt que par un manque de validité de l'effet. Cette situation illustre une limite des répliquabilités conceptuelles qui est qu'elles ne permettent pas de remettre en cause la validité d'un effet lorsqu'elles échouent. Pour cause, une répliquabilité conceptuelle implique par définition une variation notable dans la méthode. Par conséquent, si elle ne réplique pas l'effet attendu, l'échec peut être attribué à la variation méthodologique plutôt qu'à la validité de l'effet. Par exemple, si une nouvelle mesure était utilisée pour étudier l'EIC mais n'observait pas de persistance de la FI, l'explication privilégiée serait probablement d'attribuer l'absence d'effet à la nouvelle variable dépendante. La conséquence plausible dans cet exemple, selon plusieurs auteurs, serait que le manque de résultat soit jugé inintéressant, l'étude ne soit pas publiée et que la crédibilité de l'effet reste intacte (Pashler & Harris, 2012 ; Romero, 2019).

Sur le long terme, plusieurs auteurs suggèrent que ce processus pourrait mener à la persistance d'une fausse théorie ou d'un effet inexistant au sein de la littérature (Pashler & Harris, 2012 ; Romero, 2019). Les auteurs décrivent un scénario qui est qu'en raison de

pratiques scientifiques improbables (e.g., fraude) ou pour des raisons de hasard, une expérience pourrait publier une erreur de type 1 (i.e., rejet par erreur de l'hypothèse nulle) démontrant un résultat inédit. Après quoi, un autre chercheur pourrait mener une série de tentatives de répliques conceptuelles. Parmi toutes les expériences qu'il conduirait, une seule pourrait offrir des résultats significatifs. Le faux positif risquerait toutefois d'être publié et tous les vrais négatifs d'être attribués à des éléments méthodologiques et de rester méconnus. Après un certain temps, ce processus pourrait être répété par plusieurs chercheurs et contribuer ainsi à une littérature grandissante autour d'un effet pourtant fictif. En somme, bien que les répliques conceptuelles soient nécessaires pour appréhender une part de la validité d'un effet, elles ne suffisent pas à l'assurer.

Contrairement aux répliques conceptuelles, les répliques proches ont l'intérêt majeur de n'impliquer, par définition, que très peu de variations méthodologiques. Par conséquent, une réplique proche qui ne retrouverait pas les résultats initiaux d'une autre étude permettrait de questionner leur validité (Pashler & Harris, 2012). À l'inverse, une reproduction des résultats ajouterait une nouvelle preuve à leur crédibilité (Simons, 2014).

Au sein des sciences, une littérature grandissante s'est établie pour étudier la répliquabilité proche de résultats (Nosek, 2014 ; Open Science Collaboration, 2012, 2015). Beaucoup de travaux ont cependant échoué dans leurs tentatives (Bohannon, 2014), tels que ceux menés au sein du Reproducibility Project. Ce dernier a exploré la répliquabilité d'une centaine de résultats en psychologie (S. Carpenter, 2012 ; Open Science Collaboration, 2012, 2015). Parmi les 97 études qui observaient initialement un effet significatif, seules 35 ont trouvé une reproduction des résultats. Par exemple, une recherche réunissant 17 laboratoires différents (R. A. Klein et al., 2022) n'a pas répliqué le résultat selon lequel l'induction de l'idée de mortalité augmentait la tendance des individus à défendre leur vision du monde (Greenberg et al., 1994). De même, Buttrick et al. (2020) n'ont pas retrouvé le résultat selon lequel la lecture

d'un texte stipulant que le libre arbitre était une illusion augmentait la propension des participants à tricher (Vohs & Schooler, 2008). Ces divers échecs des essais de répliation proche ont mené plusieurs auteurs à évoquer une crise de la répliability au sein de la psychologie (Pashler & Harris, 2012 ; Pashler & Wagenmakers, 2012).

Le risque de faux positifs que souligne la crise de la répliability appelle à interroger la répliability proche de l'EIC. Sur ce point, la place prépondérante de l'article de Johnson et Seifert (1994) dans la littérature sur l'EIC (voir partie 2.1.2) ferait de leur travail un candidat pertinent pour évaluer la répliability proche de cet effet. Or, à notre connaissance, aucune expérience n'a directement répliqué le résultat principal de Johnson et Seifert (1994) avec succès (i.e., un nombre de références à la FI supérieur pour le groupe ayant reçu la FI avec correction par rapport à un groupe contrôle ne recevant que la correction). Quelques travaux plaident en faveur de la répliability proche de ce résultat, mais des éléments limitent la validité de cette conclusion. Par exemple, une expérience (Ecker et al., 2011) a utilisé une version du texte de Johnson et Seifert (1994) et a observé une répliation de leur résultat. Néanmoins, les auteurs ne mentionnaient pas quelles modifications avaient été apportées au texte ni si le contenu des variables dépendantes utilisées était différent, ce qui ne permet pas de déterminer à quel point leur répliation était proche du paradigme initial. De la même manière, Susmann et Wegener (2021) ont repris une adaptation du texte et des variables dépendantes utilisés par Johnson et Seifert (1994). Les résultats qu'ils observaient s'accordaient avec les prédictions théoriques de l'EIC, ce qui suggérerait l'occurrence de cet effet. Toutefois, il n'était pas possible de strictement le confirmer car leur paradigme n'incluait pas de groupe contrôle sans FI. Un dernier travail (Connor Desai & Reimers, 2019) a répondu à ces limites en incluant un groupe contrôle sans FI et en précisant les modifications apportées au paradigme initial. Les modifications étaient notamment la conduite de l'expérience sous une forme informatisée ainsi

que l'utilisation de questions inférentielles fermées. Néanmoins, dans cette expérience un échec dans la réplique du résultat de Johnson et Seifert (1994) était observé.

Au final, l'état de l'art ne permet pas de connaître directement la répliquabilité proche de l'EIC. Dans le but de perfectionner la connaissance de la validité de cet effet (Pashler & Harris, 2012 ; Romero, 2019 ; Simons, 2014), l'objectif principal de l'Expérience 2 a été d'évaluer la répliquabilité proche des résultats de Johnson et Seifert (1994). Deux groupes de participants ont été exposés au texte utilisé par les auteurs. Un groupe a reçu une version du texte comportant une FI qui était ensuite corrigée (groupe FI). Un autre groupe a lu une autre version ne comportant pas de FI (groupe contrôle). Les inférences des participants étaient évaluées par des questions fermées, adaptées de l'étude de Connor Desai et Reimers (2019).

En outre, un objectif secondaire de l'Expérience 2 a été d'évaluer la répliquabilité conceptuelle de l'EIC avec une population de participants qui n'avait à notre connaissance jamais été évaluée dans la littérature sur l'EIC. Dans la plupart des recherches sur la lutte contre les FI, les populations recrutées sont des populations américaines (Courchesne et al., 2021). Or, certains travaux supposent que la culture pourrait influencer l'EIC. Par exemple, dans une méta-analyse sur les variables modérant l'efficacité des corrections, il était observé que la taille d'effet était variable en fonction des régions du monde étudiées (Walter & Murphy, 2018). De la même manière, dans une autre étude, il était relevé qu'une intervention visant à promouvoir le consensus d'experts sur l'origine anthropogène du réchauffement climatique avait une influence variable entre une population américaine et une population australienne (Cook & Lewandowsky, 2016). Enfin, Xu et al. (2020) ont réalisé deux expériences dans lesquelles les inférences de participants étaient interrogées après la lecture d'un texte d'EIC. Les participants étaient de nationalité australienne dans la première expérience et de nationalité chinoise dans la seconde. La comparaison des résultats entre les deux expériences a mis en exergue que l'EIC était plus important pour la population chinoise. En démontrant l'influence de la culture sur

l'EIC, ces travaux interrogent le caractère généralisable de cet effet à une population francophone. C'est pourquoi un objectif secondaire de cette étude a été d'étudier la répliquabilité de l'EIC avec une population francophone, ce qui n'avait encore jamais été réalisé à notre connaissance.

3.3.1 Méthode

La recherche suit les recommandations éthiques de l'American Psychological Association et la déclaration de Helsinki. Elle s'appuie sur les recommandations de Brandt et al. (2014) dans la conduite d'une répliquabilité proche. Le matériel utilisé est accessible en ligne avec le lien suivant <https://osf.io/bcfej/>. Cette expérience a donné lieu à une publication dans la revue *Acta Psychologica* (Laurent et al., 2023).

3.3.1.1 Participants

La taille d'échantillon a été déterminée a priori en réalisant une moyenne du nombre de participants recrutés dans les cinq expériences composant l'article de Johnson et Seifert (1994) ($n = 49$). Une analyse de puissance avec le logiciel G*Power (Faul et al., 2009) a aussi été conduite en reprenant la taille d'effet large ($d = 1,521$) observée dans l'étude de Johnson et Seifert (1994, étude 1A). Elle suggérait de recruter un échantillon de huit participants par condition.

Un total de 52 participants a été recruté. Ils étaient recrutés par le biais d'annonces diffusées sur différents canaux tels que les réseaux sociaux ou par des affiches. Ces annonces étaient rédigées en français. Tous ont consenti à ce que leurs données soient analysées. La condition d'être majeur était fixée comme modalité d'inclusion. Huit participants ont été retirés des analyses pour non-respect des consignes (i.e., la somme des points qu'ils attribuaient aux questions inférentielles n'était pas constamment égale à dix). L'effectif final était alors de 44 participants (30 de genre féminin, 1 de genre autre). Parmi eux, 27 (61,363 %) avaient un niveau d'études supérieur au baccalauréat. L'âge moyen était de 23,63 ans ($\sigma = 9,49$). Les participants

étaient aléatoirement attribués au groupe FI (n = 21) ou dans le groupe contrôle (n = 23). Il n'y avait pas de différence significative d'âge ($t(42) = -0,398$, $p = 0,693$) ni dans la répartition des genres ($\chi^2 = 1,23$, $p = 0,746$) entre les groupes.

3.3.1.2 Matériel

3.3.1.2.1 Texte

Le texte utilisé était une traduction française du texte anglophone utilisé par Johnson et Seifert (1994) et Connor Desai et Reimers (2019). Il a été traduit de l'anglais puis relu et corrigé par trois chercheurs dont un bilingue expert. Le texte se composait de quinze messages écrits en noir sur fond blanc. Chaque message était présenté individuellement de manière successive sur l'écran. Le texte pouvait se décliner en une version FI ou une version contrôle. La version FI présentait une information au sixième message qui était corrigée au treizième (i.e., message 6 : « [...] *Un message de l'enquêteur de police Lucas dit qu'il y aurait eu des seaux de peinture à l'huile et des bonbonnes de gaz pressurisés dans le cagibi [...].* » ; message 13 : « [...] *Un second message de la part de l'enquêteur de police Lucas [...] indique que le cagibi qui était supposé contenir des seaux de peinture à l'huile et des bonbonnes de gaz pressurisé était en fait vide [...].* »). La version contrôle remplaçait la FI par la vraie information et la correction par une information neutre (i.e., message 6 : « [...] *le cagibi était vide [...].* » ; message 13 : « [...] *les deux pompiers qui ont été emmenés à l'hôpital viennent d'en ressortir* »). Les treize autres messages étaient les mêmes entre les deux versions.

3.3.1.2.2 Mesures

Les questions, leur modalité de réponse et leur cotation étaient basées sur le travail de Connor Desai et Reimers (2019).

Questions de rappel. Les participants répondaient à huit questions de rappel. Ils choisissaient une réponse parmi quatre qui leur étaient proposées. Les huit questions se composaient de six questions de rappel de faits généraux (e.g., « *Qu'est ce qui était présent*

dans le cagibi avant que le feu ne se déclenche ? ») et de deux questions spécifiques aux informations qui différaient entre les groupes (e.g., « *Quel était le sujet du second message reçu par la police ?* »). La bonne réponse à ces dernières variait en fonction de la version du texte (e.g., groupe FI : « *Le cagibi [...] était vide.* » ; groupe contrôle : « *Les deux pompiers qui ont été emmenés à l'hôpital viennent d'en ressortir* »). Un point était attribué à chaque réponse correcte. Le score maximum pour les questions de rappel était de huit.

Questions inférentielles. Six questions inférentielles étaient posées aux participants (e.g., « *Quelle est la cause la plus probable de l'incendie ?* »). Chaque question offrait quatre propositions de réponse dont l'une était en cohérence avec la FI (e.g., *a) Des gaz et des liquides inflammables n'ont pas été stockés de façon adéquate ; b) Quelqu'un a délibérément mis le feu aux locaux ; c) Une braise provenant de l'extérieur a mis le feu aux papiers ; d) Un mégot mal éteint se trouvait dans les poubelles*). Le participant disposait de dix points à allouer aux modalités de réponse en fonction de la probabilité qu'il leur attribuait. Plus il allouait de points, plus il jugeait cette modalité probable. Par exemple, un participant qui allouait 8 points à la proposition a), 2 points à la proposition b) et 0 point à la proposition c) et d) trouverait a) très probable, b) peu probable et évincerait les hypothèses proposées par c) et d). Un score d'EIC était mesuré en réalisant la somme des points alloués aux propositions cohérentes avec la FI (e.g., *a) Des gaz et des liquides inflammables n'ont pas été stockés de façon adéquate*). Ainsi, plus le score était haut, plus il supposait un raisonnement inférentiel en accord avec la FI.

3.3.1.3 Procédure

La procédure était informatisée et supportée par le logiciel Inquisit dans sa version web. Les participants accédaient à l'expérience depuis un lien Internet. Le lancement de la tâche commençait par une lettre d'information qui indiquait que l'étude portait sur le raisonnement envers des textes. Les participants étaient informés qu'ils auraient à lire un texte relatif à un incendie puis à répondre à des questions sur ce texte. S'ils ne s'opposaient pas à prendre part à

l'expérience, ils signalaient leur genre, leur âge et leur niveau d'études. Après quoi, une page d'instruction leur indiquait qu'ils allaient maintenant lire un texte et qu'ils auraient ensuite à répondre à des questions à son sujet. Il était demandé aux participants de lire le texte comme s'ils lisaient un article de journal et ils étaient informés qu'ils ne pourraient pas revenir en arrière dans la lecture des messages. La version du texte qui leur était ensuite présentée était aléatoirement définie par le logiciel Inquisit. Les participants lisaient le texte à leur propre rythme et passaient chaque message en appuyant sur la touche « espace ». Après la lecture, ils répondaient aux questions inférentielles puis aux questions de rappel. Ils étaient ensuite informés du but réel de l'expérience par un texte de debriefing. Il leur était alors demandé s'ils acceptaient que leurs données soient conservées. Les participants qui refusaient étaient retirés de l'analyse des données.

3.3.2 Résultats

L'analyse des résultats consistait à comparer deux groupes indépendants, à savoir les réponses des participants ayant reçu la version FI et ceux ayant reçu la version contrôle. Un test de Student était utilisé lorsque les conditions d'applications étaient réunies. Dans le cas contraire, un test U de Mann-Whitney était employé.

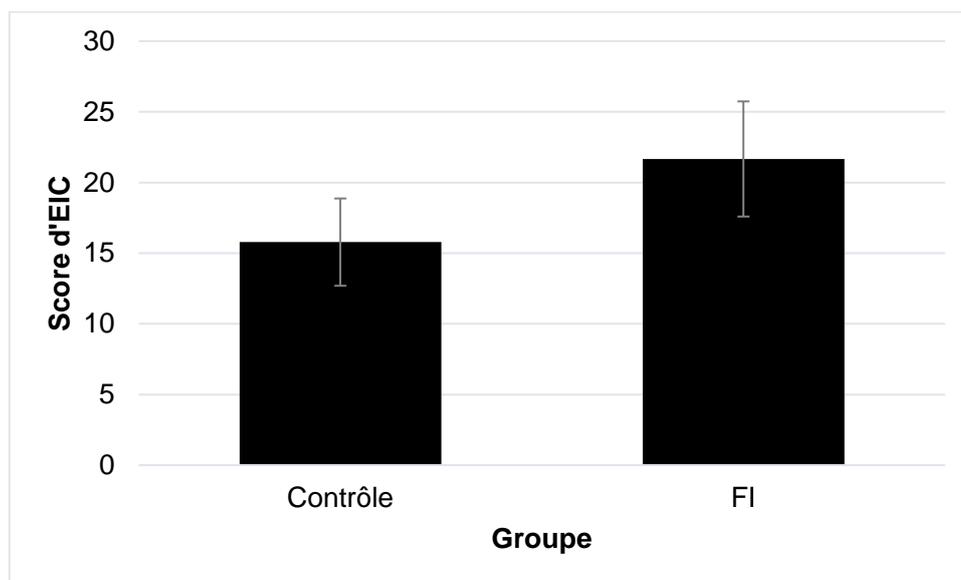
3.3.2.1 Questions de rappel

La somme des points aux questions de rappel était comparée entre les participants du groupe FI et ceux du groupe contrôle. Un test de Mann-Whitney a été utilisé en raison du non-respect de la normalité de la distribution des différences ($W = 0,937$, $p = 0,019$). Il n'y avait pas de différence significative dans les scores aux questions de rappel entre le groupe FI ($\bar{x} = 5,333$; $\sigma = 1,528$) et le groupe contrôle ($\bar{x} = 5,217$; $\sigma = 1,783$) ($U = 237$, $p = 0,924$).

3.3.2.2 Questions inférentielles

Le score d'EIC était comparé entre les participants du groupe FI et ceux du groupe contrôle. Le score obtenu par le groupe FI était significativement plus grand par rapport au

groupe contrôle ($t(42) = 2,28, p = 0,028$). En d'autres termes, les participants ayant reçu l'information fautive corrigée attribuaient plus de points de plausibilité aux propositions en lien avec la FI ($\bar{x} = 21,667 ; \sigma = 9,525$) par rapport aux participants n'y ayant pas été exposés ($\bar{x} = 15,783 ; \sigma = 7,556$). Les résultats sont présentés dans la Figure 8. La taille d'effet ($d = 0,688 ; 95 \% \text{ CI} = [0,075, 1,294]$) était modérée (Lee, 2016) et inférieure à celle observée par Johnson et Seifert (1994, étude 1A, $d = 1,521$).



Note. Les barres d'erreur sont des intervalles de confiance, $\alpha = 0,05$.

Figure 8. Score d'EIC en fonction du groupe expérimental

3.3.3 Discussion

L'objectif de cette expérience était d'étudier la réplicabilité proche du résultat de Johnson et Seifert (1994) en recrutant une population francophone. Deux groupes de participants francophones ont été exposés à une version informatisée du paradigme de Johnson et Seifert (1994). Le rappel du texte était similaire entre les groupes, suggérant que la compréhension du texte n'était pas différente en fonction de la version lue. Les résultats ont montré que les participants ayant reçu la version FI avaient plus tendance à utiliser l'information fautive pour interpréter le texte par rapport aux participants exposés à la version contrôle. Ce

résultat indique la présence d'un EIC et confirme la répliquabilité proche des résultats de Johnson et Seifert (1994) au sein d'une population francophone.

Le succès de cette répliquabilité proche soutient que le résultat de Johnson et Seifert (1994, étude 1A) n'est pas un faux positif. Cette observation constitue un apport théorique important en supportant la validité de l'EIC d'une manière inédite. De plus, dans un contexte scientifique où de nombreux résultats ne parviennent pas à être répliqués (Open Science Collaboration, 2012, 2015) et dans lequel un précédent travail sur l'EIC ne répliquait pas le résultat de Johnson et Seifert (1994) (Connor Desai & Reimers, 2019), le succès de cette répliquabilité proche a une signification particulière.

Sur le plan pratique, le support de la validité de l'EIC constitue aussi un apport majeur car il souligne qu'une correction peut être insuffisante pour évincer l'influence d'une FI. De surcroît, la présente expérience ajoute que cet effet s'applique aussi aux populations francophones, ce qui est un résultat notable dans la mesure où l'EIC est sensible à la culture (Walter & Murphy, 2018 ; Xu et al., 2020). Ainsi, le succès de cette répliquabilité proche supporte l'insuffisance possible des corrections face à l'influence de FI et montre que cette limite de l'approche corrective s'applique aussi aux pays francophones.

Une question est de déterminer pourquoi la présente expérience a répliqué avec succès un EIC tandis qu'une autre tentative de répliquabilité proche avait échoué (Connor Desai & Reimers, 2019). Des modifications mineures dans le paradigme de Connor Desai et Reimers (2019) pourraient expliquer les résultats différents observés dans leur travail. Notamment, dans la recherche actuelle, et comme dans l'étude initiale (Johnson & Seifert, 1994), les participants passaient chaque page à leur propre rythme. A l'inverse, dans l'expérience de Connor Desai et Reimers (2019), chaque message ne pouvait pas être passé avant un délai de dix secondes. Il est possible que le temps de lecture minimum imposé ait obligé les participants à analyser davantage chaque information, dont la correction (Bago et al., 2020). En accord avec l'approche

cognitive, cette manipulation a pu contribuer à la qualité de l'encodage et de l'intégration de la correction, donc à la réduction de l'EIC (Ecker, Lewandowsky, Swire, et al., 2011).

Une limite dans le succès de cette répliation proche pourrait être qu'elle n'est pas totalement semblable au protocole initial. Par exemple, les participants passaient l'expérience de façon individuelle et sur Internet plutôt qu'en groupe et sur papier. Il doit toutefois être noté que cette modification a une importance mineure dans la mesure où de précédents travaux ont observé que ces éléments ne semblaient pas empêcher l'EIC (e.g., Butler et al., 2023). De plus, étant donné l'influence d'Internet sur la problématique des FI (Bronner, 2011 ; Lazer et al., 2018), ce mode de passation semble plus écologique pour étudier l'EIC aujourd'hui. Une autre modification est que cette expérience n'a pas repris les questions ouvertes du paradigme initial mais une adaptation en questions fermées. Le recours aux questions fermées a été adopté pour limiter le risque d'abandon de tâche des participants (Connor Desai & Reimers, 2019). D'autres modifications par rapport à l'étude initiale sont que les participants n'étaient pas tous des étudiants en psychologie et n'ont pas reçu de crédits dans le cadre d'un cours. Le reste de cette expérience, telles que la base du texte, de la procédure, des mesures et des instructions étaient similaires. Étant donné que ces aspects constituent les modalités principales du paradigme, il semble acceptable d'estimer que cette recherche offre une répliation *proche* (Brandt et al., 2014) de l'étude initiale.

En conclusion, cette expérience est à notre connaissance la première à avoir répliqué de façon proche le résultat principal de Johnson et Seifert (1994). Elle montre également que cet effet et ses implications sont applicables à une population francophone. Malgré l'absence d'EIC dans l'Expérience 1, le succès de cette répliation proche couplé aux nombreuses répliations conceptuelles (voir partie 3.1) de cet effet permettent d'en attester la validité.

**Chapitre 4. Déterminants
de l'effet d'influence
continue**

En démontrant que l'influence d'une information peut persister alors même que les individus savent qu'elle est fautive, la validité de l'EIC interroge sérieusement l'efficacité et la pertinence de l'approche corrective. Certaines expériences ont notamment observé que, dans certaines conditions, une correction pouvait être sans influence sur l'adhésion à une FI (Ecker, Lewandowsky, & Apai, 2011 ; Johnson & Seifert, 1994 ; Rich et al., 2022). Par exemple, dans une étude de Johnson et Seifert (1994, étude 3A), le nombre de références aux thèmes de négligence et de matériaux inflammables était similaire pour des participants ayant reçu une correction de la FI par rapport à des participants n'ayant pas reçu de correction. Des auteurs ont alors avancé l'idée que les évidences en faveur de l'efficacité des corrections seraient mixtes (e.g., Traberg et al., 2022).

Cependant, il est important de nuancer que dans la plupart des situations observées dans la littérature, les corrections ne sont pas inefficaces mais seulement insuffisantes (voir partie 2.1.1). En effet, de nombreuses études ont observé qu'une simple rétractation permettait de réduire l'influence d'une FI (e.g., Autry & Duarte, 2021 ; Butler et al., 2022 ; Ecker & Ang, 2019 ; Miller et al., 2022 ; Porter et al., 2018 ; Porter & Wood, 2022 ; Priester et al., 1999 ; Rich et al., 2022 ; Susmann & Wegener, 2021 ; Swire-Thompson et al., 2021 ; van Huijstee et al., 2021). De même, de manière plus générale, deux méta-analyses ont étudié l'efficacité de corrections dans des domaines divers (Walter & Murphy, 2018) ou uniquement politiques (Walter et al., 2019). La taille d'effet moyenne de la correction sur l'adhésion à la FI était modérée et positive (Walter et al., 2019 ; Walter & Murphy, 2018) (pour une revue de la littérature, voir aussi Courchesne et al., 2021). Enfin, une méta-analyse sur l'EIC a comparé l'adhésion à une FI entre des groupes ayant reçu une FI puis sa correction et des groupes contrôles n'ayant pas reçu de FI. Les résultats ont montré que la différence entre ces groupes était significative mais que la taille d'effet de l'EIC était seulement de faible taille (Walter & Tukachinsky, 2020).

Dans ces méta-analyses, un point important à noter est que l'hétérogénéité des tailles d'effets était significative. En d'autres termes, l'efficacité d'une correction était généralement modérée (Walter et al., 2019 ; Walter & Murphy, 2018) et l'EIC faible (Walter & Tukachinsky, 2020). Cependant, l'intensité des effets variait en fonction des études. Cette hétérogénéité suggère que de nombreuses variables influenceraient l'efficacité d'une correction et la persistance d'une FI.

Les implications de cette hétérogénéité sont importantes pour l'approche corrective. En connaissant quelles variables impactent l'EIC, il est possible de savoir avec quelles informations, dans quelles situations et auprès de quelles personnes cet effet peut augmenter ou bien être minimisé. Or, si les déterminants de l'EIC sont clairement identifiés, alors la remise en question de l'efficacité et de la pertinence de l'approche corrective par l'EIC serait moindre, car il suffirait de mettre en place des variables propices à sa réduction pour éviter cet effet.

Le présent chapitre s'insère dans cette idée avec comme objectif de dresser une synthèse de différents déterminants connus de l'EIC. De premières parties s'attacheront à examiner les variables en lien avec la FI et la correction. Une variable particulière relative au format de la FI sera étudiée dans une expérience. Des parties suivantes étudieront les variables situationnelles et dispositionnelles en lien avec l'EIC et feront chacune l'objet d'une expérience.

4.1 Variables en lien avec la fausse information

De premières variables qui pourraient impacter l'EIC sont celles liées à la FI, telles que son caractère explicite, sa crédibilité, sa valence, sa centralité ou encore sa répétition.

4.1.1 Caractère explicite de la fausse information

L'influence d'une première variable qui pourrait être interrogée est le caractère explicite de la FI. Dans la plupart des études sur l'EIC, l'information fausse était explicitement mentionnée. Par exemple, dans l'étude de Johnson et Seifert (1994), il était explicitement

spécifié que le cagibi contenait des matériaux inflammables. De même, dans les recherches de Guillory et Geraci (2013, 2016), il était explicitement mentionné qu'un homme politique aurait reçu un pot-de-vin.

Il est toutefois important de relever que dans certains cas, une FI peut être apportée de manière implicite. Par exemple et de façon anecdotique, Seifert (2002) rapportait une série d'informations passées à la télévision. Une première mentionnait qu'une famille aurait été retrouvée morte dans son domicile le lendemain d'un dîner dans un restaurant chinois. Le lendemain, une nouvelle indiquait que l'empoisonnement alimentaire avait été exclu comme cause de la mort par le médecin légiste. Quelques jours plus tard, une dernière information circulait, indiquant que le restaurant chinois avait fait faillite. De façon plus expérimentale, seule une expérience a, à notre connaissance, étudié si le caractère implicite d'une FI modérait l'EIC. Dans une expérience que les auteurs ont répliqué, Rich et Zaragoza (2015) ont exposé leurs participants à un texte décrivant le cambriolage d'une maison. Plusieurs indices implicites dans le texte amenaient à penser que le fils de la famille était un suspect potentiel (e.g., il avait des dettes de jeu et avait la consigne de veiller sur la maison pendant leur absence). Dans la condition explicite, il était explicitement ajouté à ces indices que le fils était un suspect potentiel. Les auteurs observaient que l'EIC était plus important lorsque la FI était présentée de façon implicite par rapport à quand elle l'était de façon explicite. Une explication possible s'aligne avec l'idée qu'une cause de l'EIC serait un défaut d'activation de la FI au moment de la mise à jour du modèle mental. Puisqu'une information explicite est plus saillante et plus directement liée à la correction, il est probable que le conflit entre une FI explicite et sa correction soit plus facilement détecté, ce qui faciliterait la mise à jour du modèle mental à l'inverse d'une FI implicite (Rich & Zaragoza, 2015).

4.1.2 *Crédibilité de la fausse information*

Une autre variable qui pourrait impacter l'EIC est la crédibilité de la FI. Cette supposition s'aligne directement avec l'explication rationnelle de l'EIC selon laquelle l'influence d'une FI persisterait parce qu'elle récolte plus d'adhésion que la correction (Connor Desai et al., 2020).

Une méta-analyse qui supporte cette prédiction a comparé les tailles d'effets d'EIC entre de nombreuses études et a observé que l'EIC était plus important lorsque la crédibilité de la source de la FI était cotée comme étant haute par rapport à si elle était basse (Walter & Tukachinsky, 2020). De façon plus expérimentale, une expérience a exposé un groupe de participants à un texte d'EIC dans lequel la FI avait été préévaluée comme étant atypique (e.g., l'incendie aurait été causé par les débris d'un satellite) tandis qu'un autre recevait une FI évaluée comme étant typique (e.g., l'incendie aurait été causé par des bonbonnes de gaz). Après la correction, les résultats montraient que la persistance de la FI était plus importante lorsque la FI était typique (Rich et al., 2022, étude 2). Également, Ecker et al. (2011) ont observé qu'une manipulation supposée augmenter la crédibilité d'une FI, à savoir sa répétition (Dechêne et al., 2010), augmentait sa persistance par rapport à si elle n'avait été présentée qu'une fois. Enfin, un travail (Lewandowsky et al., 2009) a relevé que l'adhésion à une information fausse prononcée par des politiciens et qui avait été publiquement rétractée, était plus élevée chez les personnes qui déclaraient avoir confiance envers les politiciens. Etant donné qu'une FI typique, répétée et portée par une source de confiance est probablement plus crédible, ces résultats supportent l'hypothèse selon laquelle la crédibilité d'une FI augmenterait l'EIC.

4.1.3 *Valence de la fausse information*

Une autre variable qui pourrait impacter l'EIC est la valence de la FI. La plupart des études sur l'EIC utilisaient des informations fausses avec une valence négative d'intensité variable, telles que la cause d'un incendie (Johnson & Seifert, 1994), la réception d'un pot-de-

vin (Guillory & Geraci, 2013, 2016), l'idée qu'un politicien aurait collaboré avec un criminel (Thorson, 2016), que le réchauffement climatique serait un canular (Drummond et al., 2020) ou encore qu'un crash d'avion serait dû à une attaque terroriste ou aux conditions météorologiques (Ecker, Lewandowsky, & Apai, 2011). Les travaux ayant employé une FI de valence positive sont plus rares, bien que quelques exceptions existent avec des expériences ayant corrigé des informations vantant les qualités d'un magasin (Misra, 1992), d'un produit (Hamby et al., 2020), d'un hôpital (van Huijstee et al., 2021) ou encore d'une personnalité politique (Cobb et al., 2013).

Dans leur méta-analyse, Walter et Tukachinsky (2020) ont comparé la taille d'effet d'EIC lorsque la FI était codée comme étant négative par rapport à si elle était neutre ou positive. Aucune différence entre ces deux catégories n'était relevée. Toutefois, en ayant regroupé les FI neutres avec les positives, cette comparaison ne permettait pas de déterminer clairement si les FI négatives avaient un impact différent par rapport aux positives. Seuls quelques travaux expérimentaux ont comparé l'EIC en fonction de la valence de la FI. Par exemple, Guillory et Geraci (2016) ont exposé des participants à des histoires impliquant un politicien au sujet duquel une information négative (i.e., il aurait reçu un pot-de-vin) ou bien une information positive (i.e., il aurait fait une donation à une association) ou neutre était rétractée. Lorsque la FI était de valence positive ou neutre, la correction semblait pleinement efficace et aucun EIC n'était observé. A l'inverse, un EIC était relevé lorsque la FI était de valence négative. En outre, van Huijstee et al. (2021) montraient une influence de la valence de la FI à travers deux expériences au sein desquelles les participants étaient exposés à un texte d'EIC dans lequel une FI positive (e.g., un hôpital qui aurait le plus haut taux de rétablissement) ou bien une FI négative (e.g., un hôpital qui aurait le plus haut taux de mortalité) était subséquemment rétractée. Un EIC était observé lorsque la FI était négative. Toutefois, un effet de sur-correction (interprété comme étant un effet boomerang par les auteurs) était relevé dans

l'autre condition, c'est-à-dire que la perception de l'hôpital devenait plus *négative* suite à la correction de la FI positive. De la même manière, dans une autre expérience (Cobb et al., 2013, étude 2), des participants ont été exposés à une information positive sur un politicien et d'autres à une information négative. Ces informations étaient ensuite rétractées. Un EIC était observé lorsque la FI était négative, la perception du politicien devenant plus négative par rapport à avant la réception de la FI. A l'inverse, un effet de sur-correction était relevé lorsque la FI était positive, les participants développant envers le politicien une attitude plus négative par rapport à s'ils n'avaient pas reçu cette FI. Cet effet de sur-correction à la suite de la correction d'une information positive a été répliqué deux fois (Cobb et al., 2013, étude 1, 3).

En somme, les quelques travaux ayant étudié l'effet de la valence de la FI sur l'EIC ont observé qu'elle pourrait interagir avec cet effet. Les évidences suggèrent que les FI négatives seraient plus persistantes face aux corrections tandis que les FI positives seraient plus sensibles aux corrections (Guillory & Geraci, 2016) voire seraient propices à un effet de sur-correction (Cobb et al., 2013 ; van Huijstee et al., 2021).

4.1.4 Centralité de la fausse information

Une autre caractéristique des informations qui pourrait influencer l'EIC est le caractère central de la FI par rapport au reste de l'histoire. Une information est centrale lorsqu'elle a un rôle causal et qu'elle est reliée à plusieurs autres éléments du récit (Kan et al., 2021). Par exemple, dans l'expérience de Johnson et Seifert (1994), la FI corrigée selon laquelle il y avait des matériaux inflammables dans un cagibi électrique était centrale car elle permettait d'expliquer l'incendie ainsi que l'intervention des pompiers et de la police. A l'inverse, si la FI corrigée était que ce n'était finalement pas la police qui était intervenue mais la gendarmerie, son rôle aurait été périphérique puisqu'elle n'aurait que peu été liée au reste du récit.

Quelques expériences ont étudié l'effet de la centralité de la FI sur l'EIC. Dans une première (Kan et al., 2021), des participants ont été exposés à un texte d'EIC concernant

l'incendie d'une laverie. Un groupe lisait que la cause de l'incendie relevait d'une négligence de l'agent d'entretien ayant omis de nettoyer le filtre du sèche-linge (FI de centralité haute). Un autre lisait que l'incendie se serait propagé au magasin de peinture voisin (FI de centralité basse). Les participants recevaient ensuite la correction de la FI. La comparaison des résultats entre les groupes a démontré que l'EIC était plus important lorsque la FI corrigée était centrale par rapport à quand elle était périphérique. De la même manière, une autre expérience (Hamby et al., 2020) a utilisé un texte décrivant un individu qui semblait avoir un élan de lucidité après avoir bu une boisson, laissant entendre que cette dernière avait la capacité d'améliorer les fonctions cognitives. Les participants lisaient ensuite une liste d'affirmations relatives à la boisson. Un groupe de participants apprenait notamment que cette boisson était connue comme améliorant les capacités cognitives, ce qui permettait de confirmer le caractère causal de la prise de boisson sur la lucidité, et donc son caractère central dans l'histoire. Un autre groupe de participants recevait une autre information n'ayant pas de lien avec le caractère causal de la boisson. Tous apprenaient ensuite que les affirmations qu'ils venaient de lire étaient fausses. Les résultats montraient ici aussi que les participants continuaient davantage de se référer à la FI lorsqu'elle avait un caractère causal par rapport à si elle n'en n'avait pas.

Plusieurs approches théoriques permettent d'expliquer l'effet facilitateur de la centralité de la FI sur l'EIC. Dans le cadre de l'approche motivationnelle et de la réticence envers l'incomplétude des modèles mentaux, il peut être supposé qu'une information centrale tient un rôle d'autant plus important au sein du modèle mental, ce qui rendrait d'autant plus prégnante la lacune que laisserait sa correction et motiverait donc les individus à rejeter cette dernière. Dans le cadre de l'approche cognitive, une prédiction pourrait être que les informations centrales sont plus fortement encodées parce qu'elles seraient potentiellement plus saillantes. Elles seraient par conséquent davantage activées lors de la compétition d'activation avec la correction, ce qui augmenterait l'EIC.

4.1.5 Répétition de la fausse information

Dans beaucoup d'études sur l'EIC, la FI était présentée une seule et unique fois. Par exemple, dans l'étude de Johnson et Seifert (1994), la présence de matériaux inflammables n'était mentionnée qu'une fois avant d'être rétractée. Il est toutefois pertinent de relever que dans certaines situations, la FI peut être répétée. Sur le sujet de la guerre entre l'Irak et les Etats-Unis par exemple, dans les deux années suivant les attentats du 11 septembre 2001, une équipe de journalistes a identifié que le nombre d'affirmations fausses prononcées par le gouvernement américain au sujet de problèmes de sécurité que poserait l'Irak serait de 935 (Buzenberg, 2014).

La preuve la plus directe de l'impact de la répétition d'une FI sur l'EIC est portée par une étude de Ecker, Lewandowsky, Swire, et al. (2011) qui utilisait une adaptation du texte de Johnson et Seifert (1994). Pour un groupe de participants, l'information selon laquelle l'incendie avait été causé par des matériaux inflammables apparaissait trois fois. Après la correction, l'analyse des inférences montrait qu'elles étaient davantage en cohérence avec la FI par rapport à un autre groupe qui n'avait vu la FI qu'une seule fois. De façon plus générale, la méta-analyse sur l'EIC de Walter et Tukachinsky (2020) a observé une relation linéaire positive entre le nombre de répétitions de la FI et la taille d'effet de l'EIC : plus la FI était répétée, plus elle semblait persister.

Ensemble, ces travaux supportent l'idée que la répétition de la FI augmenterait l'EIC. Ce résultat est cohérent avec l'approche rationnelle. Par effet de véracité (Dechêne et al., 2010), sa multiple présentation devrait la rendre plus familière, donc plus crédible (Fenn et al., 2013 ; Pennycook et al., 2018 ; Swire et al., 2017 ; Weaver et al., 2007), et donc plus résistante aux corrections (Connor Desai et al., 2020). Ce résultat est aussi cohérent avec l'approche cognitive. Lorsqu'une FI est répétée, il est possible qu'elle soit mieux stockée en mémoire, qu'elle soit donc plus facilement récupérable et qu'elle s'active alors davantage lorsque le modèle mental

est sollicité, ce qui augmenterait l'EIC en orientant la compétition d'activation en faveur de la FI (Ecker, Lewandowsky, Swire, et al., 2011).

4.2 Expérience 3. Impact du format de la fausse information sur l'effet d'influence continue

4.2.1 Introduction

L'objectif de la partie 4.2 est d'explorer en détails l'influence d'une dernière variable en lien avec la FI qui pourrait impacter l'EIC, à savoir le format de la FI. La majorité des FI diffusées semblent être de format oral (e.g., Bernstein, 2006 ; Fox, 2019 ; Keil, 2018), ou encore de format écrit à l'image d'une publication sur un réseau social dont le texte se composerait d'éléments faux (e.g., Bode & Vraga, 2021 ; Pennycook et al., 2018).

Avec le développement de nouvelles technologies, une forme de désinformation particulière, reposant sur la manipulation d'images, s'est développée. Notons que la manipulation d'images n'est pas un phénomène nouveau. Déjà lors du 19^e siècle, la question des retouches en photographie alimentait de nombreux débats (Gunthert, 2008). De même au cours du 20^e siècle, des logiciels capables de modifier des images se développaient (Manovich, 2011). Par contre, un aspect nouveau dans la manipulation d'images est la possibilité de faire faire des choses à une personne qu'elle n'a pas faites, au sein de vidéos extrêmement réalistes nommées *deepfakes* (Westerlund, 2019).

Plusieurs travaux démontrent le caractère très réaliste des *deepfakes*. Dans une expérience (Groh et al., 2022, étude 1), des participants ont été chargés d'identifier parmi des paires de vidéos, laquelle était authentique et laquelle était un *deepfake*. Les résultats montraient que parmi les 56 paires présentées, 12 n'étaient pas correctement discriminées par 65 % des participants. De même dans une autre étude (Groh et al., 2022, étude 2), il était observé qu'environ un tiers des réponses cherchant à discriminer le caractère factice ou authentique de vidéos était incorrect.

L'émergence des *deepfakes* et leur réalisme a suscité plusieurs inquiétudes chez des auteurs (Adriani, 2019 ; Fletcher, 2018 ; Warzel, 2018 ; Westerlund, 2019). Une première inquiétude est que les *deepfakes* peuvent être utilisés en tant qu'outil de désinformation. En novembre 2020 par exemple, la première ministre néo-zélandaise de l'époque, Jacinda Ardern, apparaissait dans une vidéo prétendant la montrer en train de consommer des substances psychoactives (Reuters Fact Check, 2021 ; Thompson-Fuller, 2021). De même en 2022, au début de l'invasion russe en Ukraine, un *deepfake* dans lequel le « président ukrainien » paraissait appeler sa population à rendre les armes a été diffusé sur une chaîne nationale ukrainienne qui avait été piratée (Beurnez, 2022 ; Evon, 2022). A la même période, un *deepfake* présentait « Vladimir Poutine » exprimant dans une vidéo qu'un accord de paix avec l'Ukraine avait été trouvé (Reuters Fact Check, 2022 ; Ryckmans et al., 2022).

En plus de la désinformation, une deuxième menace avec les *deepfakes* touche au rapport qu'ont les individus envers l'authenticité des informations. Puisque les *deepfakes* seraient difficilement différenciables des vidéos authentiques (Ahmed, 2021 ; Groh et al., 2022), le risque serait que les individus développent la pensée que toutes les vidéos sont potentiellement trafiquées et qu'il serait donc rationnel de ne plus en croire aucune (Chesney & Citron, 2019). Plusieurs évidences expérimentales supportent cette capacité des *deepfakes* à augmenter l'incertitude envers l'authenticité des informations. Une expérience a par exemple montré que des participants étaient plus enclins à juger que des vidéos qui leur étaient présentées étaient des *deepfakes* quand ils avaient préalablement été avertis que les scientifiques s'inquiétaient de plus en plus de ces vidéos (Ternovski et al., 2022 ; voir aussi Barari et al., 2021). De même, une autre étude qui supporte l'effet des *deepfakes* sur l'incertitude envers l'authenticité des informations a montré que plus les participants réalisaient la tâche de discriminer des *deepfakes* par rapport à des vidéos authentiques, plus ils avaient tendance à

déclarer que les vidéos étaient falsifiées, quelle que soit leur authenticité (Groh et al., 2022, étude 2).

Selon Vaccari et Chadwick (2020), une conséquence que pourrait avoir le développement de l'incertitude envers l'authenticité des informations est une réduction de la confiance envers les sources d'information. Le raisonnement de Vaccari et Chadwick (2020) est qu'en perdant confiance dans la véracité des informations, les individus en viendraient à avoir moins confiance dans les acteurs qui les portent. Supportant cette idée, une étude a observé une corrélation positive entre la fréquence à laquelle des participants disaient avoir été exposés à des *deepfakes* sur les réseaux sociaux et leur scepticisme envers les réseaux sociaux (Ahmed, 2021). De même dans une autre recherche (Vaccari & Chadwick, 2020), certains participants ont été exposés à un *deepfake* pour lequel ils devaient juger le caractère authentique puis indiquer la confiance qu'ils avaient envers les informations vues sur les réseaux sociaux. Les résultats ont montré que l'exposition à un *deepfake* réduisait cette confiance par rapport à d'autres participants n'ayant pas vu de *deepfake* et que cette réduction était médiatisée par un sentiment d'incertitude envers l'authenticité de la vidéo.

Les constats d'une utilisation possible des *deepfakes* comme outils de désinformation et d'un effet potentiel de ces vidéos sur l'incertitude envers l'authenticité des informations nous ont amené à émettre plusieurs hypothèses relatives à l'EIC. Premièrement, étant donné que les *deepfakes* constituent des FI, il est possible qu'ils suscitent un EIC. A notre connaissance, seule une recherche portée par Vaccari et Chadwick (2020) a abordé cette question de la répliquabilité conceptuelle de l'EIC avec un *deepfake*. Des participants ont visionné un *deepfake* de « Barack Obama » prononçant “*President Trump is a totale and complete dipsh*t*” et certains apprenaient ensuite que la vidéo était falsifiée. Suggérant la présence d'un EIC, la propension des participants à affirmer que Barack Obama avait déjà insulté Donald Trump de *dipsh*t* était similaire entre les participants qui avaient reçu la correction et ceux qui ne l'avaient pas reçu.

A notre connaissance, ce résultat n'a jamais été répliqué. Par ailleurs, cette recherche n'incluait pas de groupe contrôle n'ayant pas reçu la FI, ce qui empêche de déterminer si l'absence d'effet de la correction était dû à une inefficacité de la correction ou simplement à un manque d'influence de la FI.

Une seconde hypothèse que nous émettons est que l'EIC provoqué par un *deepfake* pourrait être plus important par rapport à si la FI était textuelle. Pour cause, de précédentes expériences ont observé qu'un stimulus présenté sous la forme d'une vidéo suscitait davantage d'intérêt (Yadav et al., 2011 ; Young et al., 2018), d'attention (Alley et al., 2014) ainsi qu'une meilleure qualité de rappel (Graber, 1990 ; voir cependant Sundar, 2000) et de compréhension (voir Gyselinck, 1996) par rapport à un stimulus sous forme de texte. Ces éléments suggèrent que l'encodage et la récupération en mémoire d'un *deepfake* pourraient être plus efficaces. En accord avec l'approche cognitive, la conséquence est que la présentation d'une FI sous forme de vidéo pourrait favoriser sa récupération par rapport à une forme textuelle, ce qui pourrait favoriser l'activation de la FI lors de la compétition d'activation et donc augmenter l'EIC.

Outre cette raison cognitive, un autre argument dans la capacité des *deepfakes* à accroître l'EIC touche à la crédibilité accrue dont ils pourraient bénéficier. En effet, Sundar et al. (2021) ont observé qu'une FI sous forme vidéo était jugée plus crédible par rapport à sa retranscription écrite. Une première explication à cette crédibilité accrue des vidéos est qu'associer une image à une information permettrait de la rendre plus accessible en mémoire, donc plus familière, ce qui augmenterait son caractère vraisemblable (Fenn et al., 2013 ; Newman et al., 2012). En ce sens, plusieurs expériences ont observé que l'ajout d'un élément visuel augmentait la probabilité qu'une information soit jugée vraie même lorsque cet élément n'a pas de lien avec sa véracité (Hameleers et al., 2020 ; Sundar et al., 2021). Par exemple dans plusieurs études, la proposition « *Les noix de macadamia sont de la même famille alimentaire que les pêches.* » était plus souvent jugée vraie lorsqu'elle était accompagnée d'une

photographie de noix de macadamia par rapport à si elle ne l'était pas (Fenn et al., 2013 ; Newman et al., 2012). Une autre explication est que contrairement à un texte, une information sous forme de vidéo réduirait l'analyse des informations (Kim & Sundar, 2015), ce qui pourrait augmenter la propension à trouver crédibles des informations fausses (Bago et al., 2020 ; Bronstein et al., 2019 ; Pehlivanoglu et al., 2021 ; R. M. Ross et al., 2021). Une troisième explication à la crédibilité accrue d'un *deepfake* est que contrairement à un texte qui rapporterait une FI, un *deepfake* mettant en scène cette FI en constitue une preuve directe. Supportant cette affirmation, dans une expérience, des participants étaient plus enclins à croire qu'un individu avait fait une déclaration lorsqu'ils avaient vu la vidéo de cette individu faisant cette déclaration par rapport à s'ils avaient seulement lu sa retranscription écrite (Wittenberg et al., 2021). En accord avec l'approche rationnelle de l'EIC (Connor Desai et al., 2020), la conséquence de la crédibilité accrue des *deepfakes* par rapport aux textes est qu'ils pourraient persister davantage malgré les corrections.

Dans ce contexte, les objectifs principaux de l'Expérience 3 ont été d'évaluer 1) la répliquabilité conceptuelle de l'EIC avec un *deepfake* et 2) si l'EIC provoqué par un *deepfake* était plus important par rapport à une FI textuelle. Trois groupes de participants ont été recrutés et ont été exposés à des publications Facebook sur Jacinda Ardern (JA, première ministre de Nouvelle-Zélande au moment du recueil de données). Parmi les publications qu'ils voyaient, un groupe de participants était exposé au *deepfake* de JA (groupe *deepfake*), un autre lisait une publication Facebook qui décrivait textuellement les faits qui se produisaient dans le *deepfake* (groupe texte) et un autre recevait une information neutre (groupe contrôle). Tous les participants apprenaient ensuite l'existence du *deepfake*, ce qui devait corriger la FI pour les deux premiers groupes. Des questions inférentielles relatives à JA étaient ensuite posées.

Enfin, une autre hypothèse abordée par l'Expérience 3 est que les *deepfakes* pourraient réduire la crédibilité des sources de correction. Une expérience précédemment décrite a montré

que les *deepfakes* pouvaient augmenter l'incertitude envers l'authenticité des informations et que cette incertitude pouvait augmenter la méfiance envers les informations vues sur les réseaux sociaux (Vaccari & Chadwick, 2020). Si cet effet se généralise aux médias d'information, lesquels constituent des sources de correction majeures (Graves & Cherubini, 2016), la réduction de leur crédibilité pourrait favoriser la persistance des FI dans la société (voir partie 2.2.3). Afin d'aborder cette hypothèse, un premier objectif secondaire de l'étude suivante a été de répliquer si l'exposition à un *deepfake* pouvait augmenter l'incertitude envers l'authenticité des informations. Pour ce faire, les participants ont évalué l'authenticité de plusieurs vidéos à la fin de l'expérience. Un autre objectif secondaire a été de répliquer l'effet de l'exposition à un *deepfake* sur la méfiance envers les informations vues sur les réseaux sociaux et d'explorer si cet effet se généralisait aux médias d'information.

4.2.2 Méthodologie

Le projet de recherche a été validé par le Comité d'éthique en sciences comportementales de l'Université de Lille (référence 2022-636-S110) et a été réalisé en accord avec les standards éthiques de la Déclaration d'Helsinki de 1964. Le consentement éclairé a été obtenu pour tous les participants inclus dans cette étude. Le matériel utilisé est accessible en ligne avec le lien suivant <https://osf.io/bcfejj/>.

4.2.2.1 Population

Les conditions d'inclusion étaient d'être majeur et de ne pas être porteur de troubles du spectre de la schizophrénie. La condition relative au spectre de la schizophrénie était fixée afin de prévenir une exacerbation des difficultés de distinction entre le réel et ce qui ne l'est pas que pourrait potentiellement provoquer l'exposition à des *deepfakes*.

En cohérence avec des études précédentes sur l'EIC (e.g., Chang et al., 2019 ; Ecker et al., 2011, 2017 ; Johnson & Seifert, 1994 ; Laurent et al., 2023), nous avons décidé de recruter environ 30 participants par groupes indépendants. Un total de 68 participants a été recruté par

des annonces diffusées sur différents canaux tels que les réseaux sociaux ou par des affiches (36 de genre féminin, 2 de genre autre). Parmi eux, 46 (67,647 %) avaient un niveau d'études supérieur au baccalauréat. L'âge moyen était de 32,7 ans ($\sigma = 15,8$). Les participants étaient aléatoirement attribués au groupe *deepfake* ($n = 22$), texte ($n = 24$) ou contrôle ($n = 22$). Il n'y avait pas de différence significative d'âge ($H = 1,67$, $ddl = 2$, $p = 0,443$) ni dans la répartition des genres ($\chi^2 = 4,1$, $p = 0,365$) entre les groupes.

4.2.2.2 Matériel

4.2.2.2.1 Publications Facebook

Chaque participant visionnait un total de 34 captures d'écran de publications de médias d'information sur Facebook. Les captures d'écran ont été réalisées le 06 et le 07 juillet 2022. Elles comportaient le nom du journal, la date de publication et l'engagement suscité (i.e., nombre de commentaires, de partages et de réactions). Elles se composaient aussi d'une image (photographie ou vidéo), d'un texte au-dessus de l'image et parfois d'un titre d'article en dessous de l'image. La plupart des publications correspondaient à des captures d'écran d'authentiques publications de médias d'information sur Facebook. Lorsque les contenus étaient anglophones mais que Facebook proposait une traduction automatique, aucune modification n'était réalisée. Dans le cas où aucune traduction n'était proposée, les contenus anglophones étaient traduits en français tout en veillant à préserver la même mise en forme.

La quasi-totalité des publications étaient les mêmes entre les conditions *deepfake*, texte et contrôle. Elles abordaient des sujets relatifs à JA, tels qu'une rencontre avec le prince William, ses positions sur la lutte contre le changement climatique, l'armement, la drogue et la gestion de l'épidémie COVID-19, ou encore ses compétences politiques. La Figure 9 illustre une publication.



Figure 9. Illustration d'une publication

La seule publication qui variait entre les conditions était la quatorzième. Pour le groupe contrôle, la quatorzième publication était une vraie capture d'écran d'une publication d'un média d'information. Le caractère neutre de cette publication était assuré par son contenu (i.e., rencontre avec le Prince William) et parce que son contenu était évoqué dans une autre publication vue par tous les groupes. Pour le groupe texte, le nom du journal, la date de la publication et l'engagement étaient les mêmes que pour le groupe contrôle mais l'image et les textes étaient modifiés. L'image faisait apparaître une photographie de JA à côté d'une photographie de cannabis. Le titre de l'article et celui de la publication indiquaient que JA apparaissait dans une vidéo en train de consommer des substances psychoactives. Pour le groupe *deepfake*, le titre de la publication, le nom du journal et l'engagement suscité (partages, commentaires, réactions) étaient similaires au groupe texte. Les différences étaient qu'il n'y avait pas de titre d'article et que les photographies étaient remplacées par un *deepfake* dans

lequel « JA » apparaissait en train de consommer des substances psychoactives. Le *deepfake* utilisé a été développé par le créateur de la chaîne YouTube *Genuine Fake*. L'auteur a donné son consentement pour l'utilisation de ce deepfake dans le cadre de cette expérience. Sa version initiale de plus de deux minutes a été raccourcie à 48 secondes et a été complétée de l'image d'un bouton censé permettre de gérer le volume, en bas à droite de la vidéo, afin de rendre la vidéo semblable à une vidéo diffusée sur Facebook (voir <https://osf.io/bcfej/> pour une illustration de toutes les publications).

Une autre publication importante était la vingt-cinquième publication qui servait de correction pour les groupes *deepfake* et texte. Elle était une publication fictive créée à partir d'une capture d'écran d'une publication de l'AFP Fact-Check qui identifiait réellement le *deepfake* de JA. L'auteur, la date et les réactions de la publication initiale ont été préservés. Le titre de la publication et la photographie ont été changés pour mieux correspondre au cadre de l'étude. Cette publication était donnée dans les trois conditions. Elle est présentée dans la Figure 10.



Figure 10. Illustration de la correction

4.2.2.2.2 *Vidéos utilisées*

En plus du *deepfake* de JA, quatre vidéos d'une durée inférieure à une minute ont été utilisées. Parmi elles, deux étaient des vidéos authentiques. L'une montrait un discours de Donald Trump dans lequel il condamnait l'invasion du Capitole au lendemain de l'élection de Joe Biden et l'autre présentait un extrait de débat télévisé aux Etats-Unis. Les deux autres vidéos étaient des *deepfakes*. L'accord des auteurs pour l'utilisation de leur œuvre dans cette étude a été recueilli. Un *deepfake* a été créé par Bill Poster et Daniel Howe et prétend montrer une allocution de « Mark Zuckerberg », le PDG de Facebook, dans laquelle il semble exposer le problème du contrôle des données personnelles par une seule personne. L'autre *deepfake* a été créé par l'association Solidarité SIDA et prétend montrer un discours de « Donald Trump » dans lequel il paraît annoncer avoir trouvé un moyen d'éradiquer le SIDA.

4.2.2.2.3 Mesures

Questions inférentielles. Neuf questions inférentielles étaient posées aux participants sur lesquelles ils se positionnaient sur une échelle allant de 0 à 100. Les réponses étaient codées de sorte à ce que plus le score était haut, plus il indiquait un raisonnement en accord avec la FI. Une réponse pour laquelle un score bas impliquait une plus grande persistance de la FI était recodée en soustrayant le score brut à 100. Trois des questions étaient relatives à la perception de l'aptitude politique de JA, trois concernaient la perception de son positionnement sur la consommation de substances psychoactives et trois étaient des questions leurres sans rapport supposé avec la FI. Les résultats aux questions leurres n'ont pas été analysés au sein des données. L'alpha de Cronbach pour les trois questions sur les aptitudes politiques était satisfaisant ($\alpha = 0,815$). Concernant les questions sur la consommation de substances psychoactives, l'alpha de Cronbach était très faible ($\alpha = 0,145$). Les questions relatives à la consommation de substance sont présentées dans le Tableau 1.

Tableau 1. Codage et intitulé des questions inférentielles portant sur la consommation de substances psychoactives

Codage	Intitulé
Question 2	Certains individus accusent Jacinda Ardern de consommer régulièrement des substances psychoactives : au regard des informations dont vous disposez, à quel point jugez-vous cette proposition probable ?
Question 5	A quel degré pensez-vous que Jacinda Ardern est favorable envers la légalisation du cannabis ?
Question 8 (item inversé)	A quel point Jacinda Ardern vous paraît-elle légitime pour gérer les problématiques de drogue et d'addictions au sein de son pays ?

Confiance envers les médias et envers les réseaux sociaux. Les participants étaient invités à positionner sur une échelle allant de 0 (« Pas du tout confiance ») à 100 (« Tout à fait confiance ») la confiance qu'ils portaient envers les médias (« *A quel point avez-vous confiance*

envers les informations que vous voyez dans les médias ? ») puis envers les réseaux sociaux (« A quel point avez-vous confiance envers les informations que vous voyez sur les réseaux sociaux ? »).

Incertitude envers l'authenticité d'informations. Après le visionnage de chacune des quatre vidéos, les participants étaient invités à juger leur authenticité (e.g., « *Donald Trump a-t-il ou non condamné les attaques du Capitole ?* »). Ils pouvaient choisir entre trois modalités de réponse : « *Oui, la vidéo est authentique.* » ; « *Non, la vidéo est un deepfake.* » ; « *Je ne sais pas.* »

4.2.2.3 Procédure

La procédure était informatisée. Elle était supportée par le logiciel Inquisit dans sa version web. Des annonces diffusées sur Internet (e.g., réseaux sociaux) partageaient un lien qui permettait aux participants d'y accéder. Au démarrage de l'étude, le logiciel Inquisit assignait aléatoirement les participants à l'une des trois conditions expérimentales. L'étude débutait par la présentation d'une lettre d'information qui indiquait que l'expérience portait sur la perception des personnalités politiques sur les réseaux sociaux. Les sujets étaient informés que dans une première étape, il leur serait demandé de visionner une série de publications Facebook concernant une personnalité politique, sans contrainte de temps, comme s'ils naviguaient sur leur propre réseau social. Il leur était ensuite indiqué que, dans une deuxième étape, ils devraient visionner quatre courtes vidéos présentant des allocutions de personnes connues. Enfin, ils étaient avertis qu'après chaque étape, ils seraient invités à donner leur ressenti sur ce dont ils venaient de prendre connaissance. S'ils consentaient à participer, les participants indiquaient leur âge, leur niveau d'études et leur genre. Pour passer chaque écran, ils appuyaient sur la barre « espace » de leur clavier. La recherche se poursuivait en leur rappelant que l'objectif était de regarder une série de publications Facebook concernant une personnalité politique, comme s'ils étaient sur leur smartphone ou ordinateur. Il leur était dit

qu'aucune limite de temps ne leur était imposée mais qu'une fois une publication passée, ils ne pourraient pas revenir dessus. Un écran suivant leur indiquait que les informations qui allaient leur être présentées concernaient « Jacinda Ardern, la première ministre néo-zélandaise ». Ils étaient alors invités à débiter le visionnage des publications. Chaque publication apparaissait sur un écran différent et elles étaient toujours présentées dans le même ordre. Selon le groupe qui leur avait été attribué, à la quatorzième publication, les participants étaient exposés au *deepfake* de JA (groupe *deepfake*), à une publication énonçant les faits qui se produisaient dans ce *deepfake* (groupe texte) ou bien à une publication neutre (groupe contrôle). La vidéo du *deepfake* ne pouvait pas être passée avant sa fin. Ils lisaient ensuite les autres publications, dont la vingt-cinquième qui était censée corriger la FI pour les groupes *deepfake* et texte. Après l'exposition aux publications, les participants répondaient aux neuf questions inférentielles, suivant la séquence d'une question sur la politique, une sur la consommation de substances psychoactives, une question leurre et ainsi de suite. Les réponses aux questions inférentielles étaient suivies par l'évaluation de la confiance envers les réseaux sociaux et les médias d'information. Après quoi, il était demandé aux participants s'ils connaissaient déjà JA et s'ils avaient déjà eu connaissance de la vidéo falsifiée évoquée dans l'une des publications. Pour finir, les quatre vidéos étaient présentées. Le discours de Donald Trump sur l'invasion du Capitole était suivi du *deepfake* de Mark Zuckerberg, suivi de l'extrait du débat politique puis par le *deepfake* de Donald Trump. Les participants ne pouvaient pas passer les vidéos avant leur fin. Après chaque vidéo, l'incertitude envers l'authenticité d'informations était récoltée. Après chaque réponse, le caractère authentique ou factice du document était porté à la connaissance des participants. Après chaque *deepfake*, il leur était demandé s'ils connaissaient déjà le *deepfake* qui venait d'être présenté. Suite à cela, les participants lisaient une page de debriefing qui leur révélait le but premier de l'expérience. Ils étaient informés de l'utilité qu'auraient leurs données et il leur était laissé le choix d'accepter ou de refuser qu'elles soient

conservées à des fins de recherche. Les participants qui refusaient étaient exclus de l'analyse des données. Pour finir, ils recevaient un rappel sur le caractère factice des *deepfakes* qui leur avaient été présentés, des indices pour distinguer les vidéos authentiques des falsifiées, puis des numéros d'aide relatifs à la consommation de substances psychoactives.

4.2.3 Résultats

Des analyses paramétriques étaient utilisées lorsque les conditions d'utilisation étaient respectées et des analyses non-paramétriques dans le cas contraire.

4.2.3.1 Effet d'influence continue

La moyenne des scores aux trois questions interrogeant la perception des compétences politiques a été réalisée afin d'estimer l'effet de la condition expérimentale sur la perception des compétences politiques. En raison du non-respect de la condition de normalité ($W = 0,964$, $p = 0,050$), une ANOVA n'a pas été utilisée. Un test de Kruskal-Wallis a été employé et n'a pas relevé d'effet de la condition expérimentale sur la perception des compétences politiques ($H = 1,13$, $ddl = 2$, $p = 0,568$).

Concernant les questions sur la consommation de substances, les trois questions ont été étudiées indépendamment en raison de l'alpha de Cronbach faible ($\alpha = 0,145$). La condition de normalité n'était pas respectée pour la question 2 ($W = 0,959$, $p = 0,024$) ni pour la question 5 ($W = 0,948$, $p = 0,006$) et la question 8 ($W = 0,956$, $p = 0,018$). Des tests de Kruskal-Wallis ont donc été réalisés et aucun effet de la condition expérimentale n'a été observé sur les réponses à la question 2 ($H = 1,07$, $ddl = 2$, $p = 0,585$), la question 5 ($H = 2,09$, $ddl = 2$, $p = 0,352$) et la question 8 ($H = 0,768$, $ddl = 2$, $p = 0,681$).

Ces mêmes analyses sur la perception des compétences politiques et de la consommation de substances ont ensuite été reconduites en excluant les quatre participants ayant déclaré connaître le *deepfake* de JA. Aucune analyse n'a révélé d'effet significatif ($p > 0,1$).

Enfin, ces analyses ont été reconduites en excluant les 23 participants ayant déclaré qu'ils connaissaient déjà JA avant l'expérience. Le pattern de résultats était semblable pour la moyenne des scores aux questions inférentielles politiques ainsi qu'aux questions 5 et 8. La seule tendance observable concernait la question 2 où un effet marginalement significatif de la condition expérimentale sur les réponses était observable ($F(2,42) = 3,08, p = 0,056$). Des tests post-hoc de Tukey ont été appliqués et une différence significative était observée entre le groupe *deepfake* et le groupe texte ($t(42) = -2,47, p_{\text{tukey}} = 0,045, d = -0,968, 95 \% \text{ CI} = [-1,787 ; -0,150]$). L'adhésion à la FI semblait plus haute pour le groupe texte ($\bar{x} = 52,4, \sigma = 25,8$) par rapport au groupe *deepfake* ($\bar{x} = 27,8, \sigma = 24$). Le score du groupe contrôle ($\bar{x} = 40,7, \sigma = 25,9$) n'était pas significativement différent du groupe *deepfake* ($t(42) = 1,32, p_{\text{tukey}} = 0,392$) ni du groupe texte ($t(42) = -1,35, p_{\text{tukey}} = 0,378$)

4.2.3.2 Confiance envers les médias et les réseaux sociaux

Des analyses ANOVA ont été menées afin d'étudier si la condition expérimentale avait un effet sur la confiance des participants envers les médias et envers les informations vues sur les réseaux sociaux. Aucun effet de la condition expérimentale n'a été observé sur la confiance envers les médias ($F(2,65) = 0,159, p = 0,853$) ni sur la confiance envers les informations vues sur les réseaux sociaux ($F(2,65) = 0,434, p = 0,650$).

4.2.3.3 Incertitude envers l'authenticité des informations

Afin d'étudier si la condition expérimentale avait un effet sur l'incertitude envers les authenticités des informations, deux variables ont été créées à partir des réponses données aux quatre vidéos. Premièrement, il a été étudié si la condition expérimentale influençait la tendance à répondre « *Je ne sais pas* » à ces questions. Le nombre de réponses « *Je ne sais pas* » aux quatre vidéos a été comparé entre les groupes avec un test de Kruskal-Wallis en raison du non-respect des conditions de normalité ($W = 0,907, p < 0,001$) et d'homogénéité des variances ($F(2,65) = 2,99, p = 0,057$). Les résultats ne montraient pas d'effet de la condition expérimentale

sur le nombre de réponses incertaines ($H = 0,909$, $ddl = 2$, $p = 0,635$). Deuxièmement, il a été examiné si la condition expérimentale impactait la propension à affirmer qu'une vidéo était un *deepfake* indépendamment de sa véracité. Le nombre de réponses « *Oui, c'est un deepfake* » aux quatre vidéos a été comparé avec un test de Kruskal-Wallis en raison du non-respect de l'homogénéité des variances ($F(2,65) = 2,52$, $p = 0,089$). L'analyse ne révélait pas de différence significative ($H = 5,04$, $ddl = 2$, $p = 0,080$). Les mêmes analyses ont été conduites en excluant les quinze participants qui avaient répondu positivement à au moins une question qui interrogeait s'ils avaient déjà eu connaissance de l'un des *deepfakes* utilisés et le même pattern de résultats a été observé ($p > 0,05$).

A des fins exploratoires, il a également été examiné si la condition expérimentale impactait la capacité des participants à distinguer correctement les vraies des fausses vidéos. Un score d'identification correspondant au nombre de réponses correctes concernant l'authenticité des vidéos a été calculé. Une ANOVA a été menée en excluant les quinze participants qui avaient répondu positivement à au moins une question qui interrogeait s'ils avaient déjà eu connaissance d'un des *deepfakes* utilisés. Cette exclusion a été réalisée car la connaissance préalable du caractère factice des vidéos biaiserait la mesure de la capacité à distinguer les vraies des fausses vidéos. L'analyse n'a pas relevé d'effet de la condition expérimentale sur le score d'identification ($F(2,50) = 1,25$, $p = 0,294$).

4.2.4 Discussion

Les objectifs de cette étude étaient d'examiner la répliquabilité conceptuelle de l'EIC avec un *deepfake*, si cet effet était plus important avec un *deepfake* par rapport à une FI textuelle, et si l'exposition à un *deepfake* pouvait augmenter l'incertitude envers l'authenticité d'informations ainsi que la méfiance envers les médias d'information et les informations vues sur les réseaux sociaux. Un groupe de participants a vu un *deepfake* censé impacter leurs perceptions d'une personnalité politique, un autre était informé de l'existence du contenu de la

vidéo de façon textuelle et un autre recevait une information neutre. Tous apprenaient ensuite que la vidéo était un *deepfake*. Les participants ont ensuite répondu à des questions inférentielles évaluant leur perception de la personnalité politique, indiqué leur confiance envers les médias et les informations vues sur les réseaux sociaux, puis visionné quatre vidéos pour lesquelles ils devaient estimer l'authenticité. Les résultats ont montré que la condition expérimentale n'avait pas eu d'impact sur les réponses à ces variables.

L'absence d'EIC observée, même dans le groupe ayant reçu une FI sous forme de texte – un format pourtant bien validé pour induire cet effet –, suggère que le paradigme employé n'était pas propice à l'apparition de cet effet. Par conséquent, cette étude ne permet pas de répondre aux questions de répliquabilité conceptuelle de l'EIC par un *deepfake* ni s'il serait plus important avec cette forme de FI par rapport à une forme textuelle.

Plusieurs éléments pourraient expliquer l'absence d'EIC observable dans le paradigme employé. Une première explication pourrait être que la correction a suscité un effet boomerang dû à un effet de familiarité pour le groupe contrôle. Pour cause, les participants dans cette condition n'étaient pas exposés à la FI mais apprenaient néanmoins, dans la correction, que JA n'apparaissait pas dans une vidéo où elle aurait consommé des substances psychoactives. Ainsi, cette correction a pu rendre plus familier le lien entre JA et la consommation de substances, suscitant une augmentation de l'adhésion à la FI pour le groupe contrôle. Parallèlement, un EIC s'est peut-être produit dans les groupes *deepfake* et texte. Cependant, l'augmentation de l'adhésion à la FI dans le groupe contrôle dû à l'effet boomerang a pu empêcher de détecter l'EIC pour les deux groupes ayant reçu une FI. Cette explication doit être prise avec précaution car il n'est pas possible de démontrer que la correction a réellement eu un effet boomerang pour le groupe contrôle. Un autre groupe contrôle recevant la FI mais pas de correction aurait été nécessaire pour démontrer cet effet. De plus, rappelons que la validité de l'effet boomerang dû à un effet de familiarité est discutée au sein de la littérature (Ecker et al., 2023 ; Ecker,

Lewandowsky, et al., 2020). Il doit toutefois être noté que son occurrence reste possible, notamment dans une telle situation où les participants n'avaient probablement jamais entendu parler (Autry & Duarte, 2021) d'un lien possible entre JA et la consommation de substances (voir partie 2.3.1.3.1).

Une autre explication possible à l'absence d'EIC pourrait être que la correction a été pleinement efficace, aboutissant ainsi à une absence d'influence persistante de la FI. Supportant cette hypothèse, plusieurs éléments de la méthodologie ont pu accroître l'efficacité de la correction. Par exemple, les participants disposaient de toutes leurs ressources attentionnelles pour traiter la correction (Ecker et al., 2011). Également, toutes les informations concernaient JA, ce qui a peut-être facilité la détection du conflit entre la FI et la correction dans la mesure où les participants n'avaient pas d'autres types d'informations à traiter. Cette explication semble toutefois insuffisante étant donné que d'autres expériences ont observé un EIC malgré la présence de ces éléments méthodologiques (e.g., Johnson & Seifert, 1994).

Une troisième explication possible à l'absence d'EIC repose sur l'utilisation des mesures. La FI concernait une consommation de substances. Or, la plupart des mesures portaient plus ou moins directement sur des questions politiques. C'était le cas des questions sur les compétences politiques, mais aussi des questions 5 (favorabilité de JA envers la légalisation du cannabis) et 8 (légitimité de JA pour gérer les problématiques de drogues) relatives aux consommations de substance. Il est possible que ces mesures aient été trop éloignées de la FI pour détecter son influence. Concrètement, les participants ont pu augmenter leur adhésion à la croyance selon laquelle la personnalité consommait des substances psychoactives, sans que cette augmentation ne soit détectée par les mesures utilisées. En accord avec cette explication basée sur un manque de validité des mesures, la seule question ayant montré un effet marginalement significatif était la question 2 (probabilité que JA consomme des substances), qui est celle qui semble la plus directement liée à la FI.

Cette tendance révélait que parmi les participants qui ne connaissaient pas JA avant l'expérience, ceux ayant vu la FI sous forme de *deepfake* puis sa correction avaient significativement *moins* tendance à déclarer que JA pourrait consommer des substances par rapport à ceux qui avaient reçu la FI sous forme de texte. Le groupe contrôle ne se distinguait significativement d'aucun des deux groupes et son score tombait numériquement entre les deux. La tendance observée suggère que la présentation du *deepfake* corrigé aurait favorisé une inclination vers un effet de sur-correction tandis que la présentation de la FI sous forme de texte aurait favorisé une inclination vers un EIC. Cette tendance va à l'inverse de notre hypothèse initiale selon laquelle l'EIC serait plus important avec une FI sous forme de *deepfake* par rapport à une FI sous forme de texte. Ce résultat doit toutefois être pris avec précaution car l'effet de la condition expérimentale n'était que marginalement significatif et parce que les résultats des deux groupes avec FI ne se distinguaient pas de ceux du groupe contrôle. De plus, il s'observait seulement à condition d'exclure les participants qui connaissaient JA avant l'expérience. Si jamais ce résultat venait à être répliqué dans de futures recherches, une explication pourrait être que le conflit entre la FI et la correction a été plus explicite dans la condition *deepfake*, ce qui pourrait favoriser la baisse de l'EIC, voire la prise de conscience de l'influence de la FI et donc l'auto-correction. A l'inverse dans la condition texte, le lien entre la FI et la correction était plus implicite, ce qui pourrait faciliter l'EIC (Rich & Zaragoza, 2015).

Concernant l'incertitude envers l'authenticité des informations, aucun effet de la condition expérimentale n'a été observé. Cette absence de résultat va à l'inverse de ce que suggère la littérature théorique (Chesney & Citron, 2019 ; Meckel & Steinacker, 2021 ; Stover, 2018 ; Vaccari & Chadwick, 2020 ; Westerlund, 2019) et empirique (Barari et al., 2021 ; Groh et al., 2022 ; Ternovski et al., 2022) selon laquelle les *deepfakes* amèneraient les individus à douter de l'authenticité des informations. Il doit toutefois être noté que ces précédentes recherches utilisaient des avertissements explicites sur l'existence des *deepfakes* (Barari et al.,

2021 ; Ternovski et al., 2022) ou bien de multiples vidéos (Groh et al., 2022). Dans le cadre de cette expérience, la seule différence entre le groupe *deepfake* et les autres groupes était le visionnage d'un *deepfake*. Une première explication à l'absence d'effet de la condition expérimentale sur l'incertitude envers l'authenticité des informations est alors que la présentation d'un unique *deepfake* a été insuffisante pour l'augmenter. La présentation de plusieurs *deepfakes* ou bien la répétition du même *deepfake* seraient probablement davantage propices à susciter un développement de l'incertitude envers les informations. Une autre explication réside dans le fait que, dans la correction, les groupes texte et contrôle ont également reçu une information mentionnant un *deepfake* circulant sur Internet. La correction a donc pu augmenter l'incertitude envers l'authenticité des informations pour les groupes texte et contrôle, aboutissant à une absence de différence entre les trois conditions expérimentales.

L'absence d'effet de la condition expérimentale sur l'incertitude envers les informations permet d'expliquer son absence d'effet sur la méfiance envers les réseaux sociaux et les médias d'information. Pour cause, selon les résultats de Vaccari et Chadwick (2020), l'effet de l'exposition à un *deepfake* sur la méfiance envers les sources d'informations serait médiatisé par une hausse de l'incertitude envers l'authenticité des informations. Etant donné qu'une hausse de l'incertitude envers l'authenticité des informations n'a pas été observée, il est compréhensible que la condition expérimentale n'ait pas eu d'impact sur la confiance envers les médias d'information et les informations vues sur les réseaux sociaux.

Ainsi, cette expérience n'a pas observé d'EIC ni d'effet de la condition expérimentale sur l'incertitude envers l'authenticité des informations, ce qui ne permet pas d'explorer si un EIC est plus important lorsqu'il est amorcé par un *deepfake* ni si l'exposition à un *deepfake* augmente la méfiance envers certaines sources ou canaux d'informations. De futures recherches pourraient chercher à réévaluer ces hypothèses en évitant plusieurs limites du paradigme utilisé, comme en employant des mesures plus valides et un groupe contrôle avec FI mais sans

correction. Un autre ajustement pourrait être de présenter plusieurs fois la FI pour le groupe texte et *deepfake*, ce qui permettrait d'exposer les participants du groupe *deepfake* à plusieurs *deepfakes*. Cet ajustement offrirait probablement une situation plus adaptée pour évaluer l'influence des *deepfakes* sur l'incertitude envers l'authenticité des informations et la méfiance envers les sources d'information.

Par ailleurs, bien que la ressemblance avec la manière dont les informations sont traitées sur les réseaux sociaux ait été recherchée par le biais des instructions et de l'utilisation de publications Facebook, plusieurs éléments méthodologiques limitent le caractère écologique de ce paradigme. Premièrement, les participants disposaient de leurs pleines ressources attentionnelles. Or, sur les réseaux sociaux, il est probable que la disponibilité des ressources attentionnelles soit différente. Une étude conduite en 2021 a demandé à un échantillon à quelle fréquence ils se retrouvaient en multi-tâche lorsqu'ils étaient sur les réseaux sociaux (e.g., écouter une vidéo tout en utilisant Facebook) et la moyenne des réponses scorait entre « parfois » et « souvent » (Jeong & Hwang, 2012). Considérant que la disponibilité des ressources attentionnelles impacterait l'EIC (Ecker, Lewandowsky, Swire, et al., 2011), il est possible que cette modalité ait influencé les résultats. Une autre limite au caractère écologique est que chaque information était présentée sur un écran différent que passaient les participants en appuyant sur une touche de leur clavier. Or, dans des conditions réelles sur les réseaux sociaux, les individus feraient défiler les différentes informations sur un unique écran. Ces conditions seraient peut-être propices à une analyse plus rapide des informations, ce qui pourrait réduire l'impact de la correction si elle est insuffisamment intégrée. Enfin, une dernière limite au caractère écologique de ce paradigme est que toutes les publications concernaient JA. Or, sur leur propre réseau social, les individus pourraient voir des informations sur JA, mais ces informations seraient mélangées à d'autres contenus, tels que des photos de vacances de leurs amis ou des publications de leurs pages favorites. Ici aussi, cette modalité a pu influencer l'EIC

car une diversité d'informations aurait pu rendre plus difficile la détection du conflit entre la FI et la correction. Les futures recherches s'intéressant à l'effet d'un *deepfake* sur les réseaux sociaux pourront veiller à éviter ces limites par des ajustements simples, tels que l'ajout d'une seconde tâche, l'utilisation d'une unique page à faire défiler contenant toutes les informations ou encore en mélangeant les informations d'intérêt avec des informations d'un contenu différent.

Pour conclure, cette recherche n'a pas mis en évidence d'EIC ni d'impact de la condition expérimentale sur l'incertitude envers l'authenticité des informations, ce qui est explicable par plusieurs éléments méthodologiques. Cette absence d'effet n'a pas permis d'étudier des hypothèses complémentaires, telles qu'une hausse de la méfiance envers les sources d'informations à la suite de l'exposition à un *deepfake* ainsi qu'un EIC plus élevé lorsqu'il est amorcé par un *deepfake*. La seule tendance observée allait plutôt dans le sens inverse de cette dernière hypothèse, mais elle resterait à confirmer par de futures recherches. Finalement, de futurs travaux restent nécessaires pour mieux comprendre l'influence du format de la FI sur l'EIC.

4.3 Variables en lien avec la correction

Outre les variables en lien avec la FI, d'autres variables qui pourraient impacter l'EIC sont celles en lien avec la correction, telles que sa formulation, son format, sa crédibilité, sa répétition ou encore sa cohérence avec les attitudes initiales de la personne qui la reçoit.

4.3.1 Formulation de la correction

De façon générale, une correction consiste à indiquer qu'une information est fausse. Dans les faits, cette correction peut être formulée de manière plus ou moins élaborée. A partir de la littérature, nous proposons de définir deux formulations de correction : une rétractation ou une réfutation.

Nous définissons une rétractation comme étant la formulation d'une correction brève qui se contente de nier la FI. Une rétractation peut se présenter de trois manières différentes. Une première consiste à ajouter l'étiquette de faux à la FI. Par exemple, Skurnik et al. (2005) ont exposé leurs participants à une série d'affirmations, dont la véracité était révélée par un mot (« vrai » ou « faux ») apparaissant sur l'écran immédiatement après leur présentation. De même, Ecker, Lewandowsky, et al. (2020) ont présenté à leurs participants différentes informations dont certaines étaient fausses. Une FI concernait par exemple le taux de crimes commis sur des personnes de couleur de peau blanche par des personnes de couleur de peau noire. Cette information était corrigée en la juxtaposant à une croix rouge et le mot « FAUX ». Outre l'ajout d'une étiquette de faux, un autre type de rétractation consiste à rétracter la FI par une phrase, sans mentionner explicitement la FI. Par exemple, dans une expérience (Kan et al., 2021), des participants ont été exposés à un texte d'EIC concernant un incendie qui était initialement attribué à un filtre de sèche-linge obstrué. Dans une condition expérimentale, cette information était rétractée en indiquant uniquement qu'un journaliste avait corrigé l'information initiale concernant la propagation du feu. Enfin, un dernier type de rétractation consiste à rétracter la FI en la mentionnant. Par exemple, dans une expérience de Johnson et Seifert (1994, étude 1B), la correction disait qu'il n'y avait finalement pas de matériaux inflammables dans le cagibi qui était censé les contenir et que ce dernier était vide. De même, dans une condition expérimentale de l'expérience de Kan et al. (2021), la FI était corrigée en indiquant qu'un journaliste avait corrigé l'information initiale concernant la propagation du feu et qu'il n'était pas dû à un filtre de sèche-linge obstrué (voir aussi Ecker et al., 2023 ; Swire et al., 2017).

L'autre formulation de correction est une réfutation, que nous concevons comme étant la correction détaillée d'une information, qui explique pourquoi l'information est fautive et/ou lui apporte une explication alternative. Deux types de réfutation sont fréquemment utilisés dans

la littérature. Un premier consiste à rappeler les faits en même temps que la FI. Par exemple, une expérience (Ecker et al., 2023) a employé des messages correctifs relatifs aux FI sur le COVID-19 en juxtaposant les mythes aux faits. De même, dans une autre expérience (Ecker, O'Reilly, et al., 2020), une correction reprenait une FI indiquant que 79 % des victimes de couleur de peau blanche étaient tuées par des personnes de couleur de peau noir, en indiquant textuellement qu'elle était fausse et en précisant la vraie statistique. Le second type de réfutation sont les corrections qui apportent une explication alternative à la FI. Par exemple, dans une condition expérimentale de l'expérience de Kan et al. (2021), un groupe de participants apprenait que l'incendie n'était finalement pas dû à un sèche-linge obstrué mais à un câblage électrique défectueux. Également, dans une condition expérimentale de l'étude de Johnson et Seifert (1994, étude 3A), les participants apprenaient que l'incendie n'était finalement pas dû à des matériaux inflammables mais que des matériaux servant à un incendie criminel avaient été retrouvés. Les différentes formulations de correction sont synthétisées dans le Tableau 2.

Tableau 2. Synthèse des formulations, types de correction et exemples.

Formulation	Type	Exemple
Rétractation	Etiquette de faux	Le cagibi contenait des matériaux inflammables. FAUX
	Phrase sans mention de la FI	Le cagibi était vide.
	Phrase avec mention de la FI	Le cagibi qui était supposé contenir des matériaux inflammables était vide.
Réfutation	Rappel des faits	Des matériaux inflammables étaient censés se trouver dans le cagibi. Or, le cagibi était vide. Personne n'avait l'accès à cette pièce.
	Explication alternative	L'incendie n'était pas dû à des matériaux inflammables. Des matériaux servant à un incendie volontaire ont été retrouvés.

Note. Les exemples sont adaptés de l'étude de Johnson et Seifert (1994).

De nombreux travaux suggèrent que les réfutations seraient plus efficaces pour réduire l'adhésion à une FI par rapport aux rétractations. Concernant les réfutations juxtaposant les mythes aux faits, Ecker, O'Reilly, et al. (2020) montraient que l'information concernant le taux de crimes réalisés par des personnes de couleur de peau noire était mieux corrigée par une réfutation que par une étiquette négative. Également, dans l'évaluation des messages correctifs relatifs à des FI sur le COVID-19, une correction rappelant les faits réduisait davantage l'adhésion par rapport à une rétractation ne mentionnant que le mythe (Ecker et al., 2023). Il doit cependant être noté que l'avantage de juxtaposer les mythes aux faits n'a pas été clairement répliqué dans une recherche (Swire-Thompson et al., 2021). Concernant le second type de réfutation, deux méta-analyses conduites sur l'EIC (Walter & Tukachinsky, 2020) ou sur les variables modérant l'efficacité des corrections (Walter & Murphy, 2018) ont observé que l'adhésion à la FI réduisait davantage lorsque les corrections fournissaient une explication alternative. Par exemple, dans leurs expériences (Johnson & Seifert, 1994 ; Kan et al., 2021), les auteurs relevaient que l'ajout d'une explication alternative permettait de réduire davantage l'EIC par rapport à s'il n'y en n'avait pas. Ce résultat a été répliqué par de nombreuses expériences (Autry & Duarte, 2021 ; Ecker et al., 2010 ; Ecker, Lewandowsky, & Apai, 2011 ; Jin et al., 2022 ; O'Rear & Radvansky, 2020 ; Rich et al., 2022 ; Swire et al., 2017).

En accord avec l'approche motivationnelle, une explication possible à l'efficacité des réfutations est que l'ajout d'une explication alternative ou le rappel des faits permettrait de mieux combler la lacune au sein du modèle mental par rapport à une rétractation, ce qui rendrait moins inconfortable l'intégration de la correction. Une explication complémentaire d'ordre cognitive serait que l'ajout de détails en lien avec la correction permettrait de la récupérer plus facilement lorsque le modèle mental est sollicité, ce qui faciliterait alors l'activation de la correction au moment de la compétition d'activation.

Enfin, il doit être souligné que l'efficacité d'une rétractation mentionnant la FI a parfois suscité des discussions. Pour rappel (voir partie 2.3.1.2.2), certains auteurs (e.g., Skurnik et al., 2005) se sont inquiétés que la répétition de la FI au sein de la correction la rende plus familière (Dechêne et al., 2010) et suscite alors un effet boomerang. Il est toutefois important de rappeler que de nombreux travaux vont à l'encontre de cette prédiction (e.g., Walter & Tukachinsky, 2020) et suggèrent parfois qu'une rétractation avec mention de la FI pourrait être plus efficace par rapport à si elle ne la mentionnait pas. Dans leur expérience par exemple, Kan et al. (2021) montraient que l'EIC était moins élevé lorsque la FI était mentionnée dans la correction (i.e., l'incendie n'était pas dû à un filtre de sèche-linge) par rapport à quand elle ne l'était pas (pour des répliques, voir Ecker et al., 2017 ; Johnson & Seifert, 1994, études 1A et 1B ; Xu et al., 2020).

En somme, il existe de nombreuses manières de formuler une correction. La manière la plus efficace serait d'ajouter une explication alternative ou d'expliquer pourquoi l'information est fautive. Une autre façon, plus rapide mais moins efficace, serait de simplement nier l'information, en lui ajoutant une étiquette de faux ou par une phrase mentionnant ou ne mentionnant pas la FI. Dans tous les cas et allant à l'encontre de précédentes inquiétudes au sein de la littérature, aucune formulation ne semble plus propice qu'une autre à un effet boomerang.

4.3.2 *Format de la correction*

De nombreuses expériences sur l'EIC ont utilisé des corrections sous format textuel (e.g., Ecker et al., 2020 ; Johnson & Seifert, 1994). D'autres formats sont toutefois envisageables. Premièrement, parmi les organisations de vérification des faits, beaucoup ont décidé d'ajouter à leurs textes des échelles permettant d'illustrer rapidement l'authenticité des informations telles que des échelles de vérité (voir Figure 11). Environ 80 % des organisations de vérification des faits useraient de telles échelles (Graves & Cherubini, 2016 ; Stencel et al.,

2022). Par exemple, le site Politifact (<https://politifact.com>) classe les informations sur une échelle de vérité nommée *Truth-O-Meter*, allant de « vrai » à « mensonge éhonté » (*pants on fire*) en passant par les qualificatifs « presque vrai », « moitié vrai », « presque faux » ou « faux ». Une expérience a étudié si l'ajout d'une échelle de vérité à une correction sous format textuel permettait de réduire l'EIC. Des participants ont été exposés à une FI politique ou bien non-politique. Cette information était ensuite corrigée par un article de vérification des faits qui pouvait contenir ou non une échelle de vérité. Les résultats ont montré que la correction de FI non-politiques était plus efficace lorsqu'elle incluait une échelle de vérité, mais ce résultat n'était pas répliqué pour les FI politiques (Amazeen et al., 2018). Par ailleurs, une méta-analyse conduite sur l'efficacité de corrections de FI politiques trouvait des tailles d'effet similaires lorsque l'article de vérification contenait une échelle ($d = 0,28$) ou n'en contenait pas ($d = 0,32$) (Walter et al., 2019).

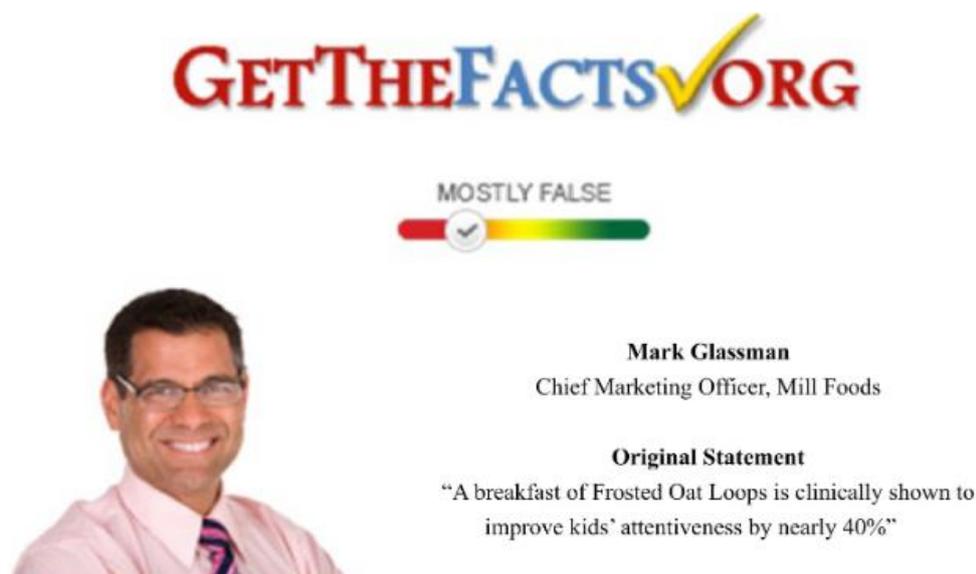


Figure 11. Exemple d'une correction incluant une échelle de vérité. Capture d'écran de l'article de Amazeen et al., 2018

Outre les échelles de vérité, une autre modification possible du format de la correction est d'utiliser une image. Dans une expérience (Hameleers et al., 2020), des participants ont lu une information accusant un individu d'être un terroriste. Ils apprenaient ensuite que l'individu avait lancé une procédure en justice contre ces allégations. Dans une condition expérimentale, les participants voyaient, en plus de cette correction sous forme de texte, une image de l'individu siégeant au tribunal. Les résultats montraient que la réduction de l'adhésion à la FI était similaire entre les groupes. Enfin, dans une autre expérience (van der Linden et al., 2014), des participants ont lu un message qui mettait en avant le consensus scientifique sur le sujet du changement climatique et ce message était accompagné d'une image dans une condition expérimentale. Ici aussi, les résultats ne montraient pas de différence d'efficacité corrective lorsqu'une image était présentée.

Une dernière modification au format de la correction peut être d'utiliser une vidéo. Dans une expérience (Young et al., 2018), après avoir été exposé à une information fautive, un groupe de participants a vu un article de vérification des faits textuel tandis qu'un autre était exposé à une adaptation vidéo de cet article. Les résultats montraient que la correction vidéo était plus efficace que la textuelle. Une conclusion similaire était réalisée dans une autre recherche (Goldberg et al., 2019) qui observait qu'une vidéo était plus efficace pour augmenter l'adhésion au consensus scientifique sur le changement climatique par rapport à une retranscription écrite de cette vidéo.

En somme, une correction peut prendre différents formats avec l'ajout d'éléments visuels. L'EIC ne semble pas sensible à l'ajout d'une échelle de vérité ou d'une image à la correction, mais il semble être davantage réduit lorsque la correction est sous format vidéo. Des explications à l'effet du format vidéo d'une information sur l'EIC ont été proposées dans la partie 4.2.1.

4.3.3 *Crédibilité de la correction*

Une autre caractéristique qui peut varier lorsqu'une FI est corrigée est la crédibilité de la correction. Des analyses conduites après la lecture de textes d'EIC ont effectivement montré que la croyance dans la correction prédisait négativement les inférences qui étaient réalisées en lien avec la FI (Ecker & Antonio, 2021). De même, une autre recherche a directement demandé aux participants s'ils croyaient ou non en la correction. Lorsque les sujets exprimaient qu'ils adhéraient à la correction, l'influence de la FI était réduite. À l'inverse, ceux qui déclaraient ne pas croire en la correction continuaient de se référer autant à l'information fautive par rapport à un autre groupe n'ayant pas reçu de correction (O'Rear & Radvansky, 2020).

De façon plus expérimentale, plusieurs travaux semblent montrer un effet de la manipulation de la crédibilité de la correction sur l'EIC. Une première manière indirecte de manipuler la crédibilité de la correction peut consister à manipuler son support social sur les réseaux sociaux. En effet, une expérience a observé que le nombre de *likes* recueillis par une information sur Twitter prédisait sa crédibilité (Butler et al., 2023). Or, il est également relevé qu'un nombre accru de *likes* sur un tweet de *fact-checking* augmentait son efficacité (Butler et al., 2023). De même, une autre expérience a observé que les participants réduisaient davantage leur croyance en une FI si la correction était fortement supportée socialement (i.e., entre 10 et 30 milliers de retweets, likes et commentaires) par rapport à si elle l'était faiblement (i.e., entre 10 et 30 retweets, likes et commentaires) (Vlasceanu & Coman, 2022).

Outre le support social, une autre manière indirecte de manipuler la crédibilité d'une correction peut être de lui attribuer des caractéristiques scientifiques étant donné que les individus ont généralement confiance en la science (CEVIPOF, 2024). Or, plusieurs études ont observé que les participants réduisaient davantage leur adhésion à une FI si la correction était formulée comme étant un fait scientifique par rapport à une anecdote (Vlasceanu & Coman,

2022), ou lorsque la correction était accompagnée d'une source par rapport à si elle ne l'était pas (Vraga & Bode, 2017a).

Les évidences supportant l'effet de la crédibilité de la correction sur l'EIC par une hausse du support social ou une attribution de caractéristiques scientifiques restent toutefois indirectes, dans la mesure où la crédibilité de la correction n'était pas directement contrôlée. Une troisième façon plus directe de manipuler la crédibilité de la correction consiste à faire varier celle de sa source. Bien que la méta-analyse de Walter et Tukachinsky (2020) n'ait pas relevé de différence de taille d'effet de l'EIC lorsque la crédibilité de la source était codée comme étant haute ou basse, plusieurs travaux l'ayant manipulé de façon expérimentale ont observé qu'elle aurait un impact sur l'EIC. Par exemple, une recherche a relevé que la correction d'une FI relative à un virus était plus efficace lorsqu'elle était fournie par une institution de santé par rapport à un simple utilisateur (Vraga et Bode, 2017). De même, la correction d'une FI concernant un politicien était plus efficace lorsque la fiabilité (i.e., la perception de sa volonté pour produire des affirmations correctes) et l'expertise (i.e., la perception de sa capacité à réaliser des affirmations correctes) de la source de correction avaient été validées comme étant hautes (e.g., un rapport gouvernemental) par rapport à quand elles étaient basses (e.g., un acteur célèbre) (Guillory & Geraci, 2013, étude 1).

Quelques recherches (Ecker & Antonio, 2021 ; Guillory & Geraci, 2013) ont d'ailleurs étudié de manière plus spécifique si la fiabilité et l'expertise de la source de la correction produisaient des effets distincts sur l'EIC. Dans une expérience (Ecker & Antonio, 2021), des participants ont lu plusieurs textes d'EIC au sein desquels la fiabilité et l'expertise de la source étaient indépendamment manipulées. Par exemple dans un texte, une FI décrivant une eau polluée par une industrie était corrigée par 1) un représentant de la pêche local (haute expertise, haute fiabilité) ou bien 2) un ingénieur en environnement de l'industrie (haute expertise, faible fiabilité), 3) un membre d'un groupe de conservation (faible expertise, haute fiabilité) 4) ou

encore un employé de l'industrie (faible expertise, faible fiabilité). Les résultats montraient que les participants se référaient moins à la FI lorsque la source de la correction avait été validée comme ayant une fiabilité haute. Par contre, cet effet inhibiteur de la crédibilité de la correction sur l'EIC ne s'observait pas lorsque seule l'expertise de la source était haute. Ce pattern de résultats a également été observé dans une autre étude (Guillory & Geraci, 2013). En somme, la crédibilité de la source réduirait l'EIC, ce qui s'accorde avec l'approche rationnelle, et cette réduction serait davantage attribuable à la caractéristique de fiabilité de la source qu'à celle d'expertise.

4.3.4 Répétition de la correction

Une autre variable qui semble impacter l'EIC est la répétition de la correction. Comme mentionné précédemment (voir partie 2.3.1.3.1), plusieurs expériences ont observé que la répétition d'une correction permettait à cette dernière d'être plus efficace (Ecker, Lewandowsky, Swire, et al., 2011). Par exemple, dans une expérience (Carnahan et al., 2021) des participants ont lu un article de *fact-checking* puis ont évalué leur adhésion à l'information qui avait été corrigée. Pour certains participants, un article de vérification des faits portant sur le même sujet avait déjà été présenté deux semaines auparavant. Supportant l'impact de l'efficacité de la répétition de la correction, les sujets pour lesquels une correction avait été déjà été présentée montraient une plus grande réduction de l'adhésion à la FI. Par ailleurs, notons que l'influence de la répétition de la correction sur l'EIC s'observe même quand la correction mentionne la FI (Sanderson et al., 2022), ce qui va à l'encontre des inquiétudes (Guess & Coppock, 2020 ; Swire-Thompson et al., 2020) qu'avaient suscitées des travaux autour de l'effet boomerang dû à un effet de familiarité (Pluviano et al., 2017 ; Skurnik et al., 2005).

Une explication possible à l'effet inhibiteur de la répétition des corrections sur l'EIC est d'ordre rationnel puisqu'une affirmation répétée serait plus crédible (Dechêne et al., 2010). Cet effet est aussi explicable avec l'approche cognitive. En permettant plusieurs expositions à

la correction, la répétition pourrait augmenter la probabilité que la correction soit encodée et intégrée au sein du modèle mental. De plus, la répétition pourrait rendre la correction plus disponible en mémoire et donc faciliter sa récupération lors de la compétition d'activation.

Quelques travaux nuancent toutefois l'efficacité de la répétition de la correction. Dans une recherche (Ecker, Lewandowsky, Swire, et al., 2011), des participants ont lu un texte d'EIC dans lequel soit la FI, soit la correction, soit ces deux informations étaient répétées. Lorsque la FI avait également été répétée, les résultats montraient que la répétition de la correction permettait de réduire davantage l'adhésion à la FI par rapport à une seule présentation. Toutefois, l'effet inhibiteur de la répétition de la correction sur l'EIC n'était plus observé lorsque la FI n'avait pas été répétée. Également, Vraga et Bode (2017) relevaient que la correction d'une FI de santé était plus efficace si elle était donnée deux fois, par un institut de santé et par un internaute, par rapport à si elle n'était donnée qu'une seule fois par un internaute. Toutefois, sa répétition ne la rendait pas plus efficace lorsqu'elle était comparée à un groupe ayant reçu une seule fois la correction par un institut de santé. Ces résultats confirment la validité de l'effet inhibiteur de la répétition de la correction sur l'EIC mais suggèrent en même temps que cette interaction dépendrait d'autres variables.

4.3.5 Cohérence entre la correction et les attitudes initiales

Enfin, une autre variable en lien avec la correction qui pourrait impacter l'EIC est la cohérence entre le contenu de correction et les attitudes initiales. Certaines informations, dites pro-attitudinales, se définissent par le fait d'aller dans le sens de ce que pense un individu. A l'inverse, les informations contre-attitudinales vont à l'encontre des attitudes initiales. Par exemple, aux Etats-Unis, une information supportant le parti démocrate serait une information pro-attitudinale pour les individus démocrates et contre-attitudinale pour les individus républicains. De la même manière, la correction d'une information pro-démocrate serait une

correction contre-attitudinale pour les individus démocrates et pro-attitudinale pour les républicains.

Plusieurs études suggèrent que le caractère pro- ou contre-attitudinale d'une correction pourrait interagir avec l'EIC. Par exemple, une étude a observé que lors du mandat de Barack Obama, les individus qui continuaient de croire que le président n'était pas né aux Etats-Unis malgré la diffusion publique de son certificat de naissance n'étaient pour la plupart pas démocrates (Tuchman, 2011). De façon plus expérimentale, plusieurs recherches ont relevé un effet d'interaction entre les attitudes initiales et l'effet de la correction sur l'adhésion à une information fausse. Il a par exemple été observé qu'un message mettant en avant le consensus scientifique sur l'origine anthropogène du changement climatique augmentait l'adhésion au consensus, sauf chez certains participants se déclarant républicains (Bolsen & Druckman, 2018 ; Cook & Lewandowsky, 2016). De même dans une autre expérience, la correction d'une FI pro-démocrate était efficace chez les participants républicains mais pas chez les démocrates (voir aussi Nyhan et al., 2020 ; Nyhan & Reifler, 2010). De façon plus générale, deux méta-analyses ont relevé qu'une FI était moins bien corrigée lorsque la correction était contre-attitudinale (Walter et al., 2019 ; Walter & Tukachinsky, 2020).

L'effet d'interaction entre l'attitude initiale et le caractère pro- ou contre-attitudinale de la correction sur l'EIC est en cohérence avec l'approche motivationnelle basée sur le raisonnement motivé. Ainsi, un individu aurait plus de probabilités d'être motivé à rejeter une correction si elle contredit une information qui s'accorde avec une attitude importante.

Il est toutefois important de nuancer que dans certains cas, le caractère contre-attitudinale d'une correction n'augmente pas l'EIC. Ecker, Lewandowsky, et al. (2014) ont conduit une étude avec des participants basés en Australie. Les participants ont lu un texte d'EIC dans lequel un individu aborigène était initialement accusé d'un cambriolage avant d'apprendre que le suspect n'appartenait finalement pas à cette population. Contrairement à ce

que prédit l'effet facilitateur d'une correction contre-attitudinale sur l'EIC, les résultats montraient que la correction était aussi efficace chez les participants qui avaient des attitudes initiales négatives envers les individus aborigènes par rapport à ceux qui n'en avaient pas. De même dans une autre expérience (Ecker & Ang, 2019), une condition expérimentale a exposé des participants à un texte d'EIC dans lequel un politicien de gauche était initialement accusé de détournement de fonds puis ne l'était plus. Ici aussi, même si les participants de droite faisaient plus de références à la FI (effet principal de l'orientation politique), la réduction de l'adhésion à la FI était aussi importante chez les participants de gauche que chez ceux de droite (voir aussi Thorson, 2016, étude 2).

Une explication proposée par Ecker, Lewandowsky, et al. (2014) à cette absence d'effet d'interaction est que la correction d'une FI ne susciterait pas de motivation à rejeter la correction lorsque la FI a un caractère spécifique. Dans les précédentes expériences (Ecker, Lewandowsky, et al., 2014 ; Ecker & Ang, 2019 ; Thorson, 2016), les FI ne concernaient qu'un individu aborigène ou un seul politicien, non pas leur groupe entier. Or, accepter qu'un aborigène particulier ou qu'un politicien du parti opposé n'est pas un criminel n'impliquerait pas de remettre en cause la croyance que la plupart des aborigènes ou des politiciens du parti opposé le sont. Par conséquent, la correction contre-attitudinale d'une FI spécifique ne remettrait pas suffisamment en question l'attitude initiale pour motiver le rejet de la correction, ce qui expliquerait l'absence d'effet d'interaction observée dans ces recherches.

Le corollaire de cette explication est que l'effet d'interaction entre les attitudes initiales et la correction sur l'EIC devrait par contre s'observer lorsque les FI ont un caractère général. Cette prédiction a été évaluée dans une autre condition expérimentale de l'étude de Ecker et Ang (2019, étude 1). Dans cette condition, des participants ont lu un texte dans lequel une information subséquentement rétractée indiquait que (tous) les politiciens de gauche étaient plus enclins à détourner des fonds par rapport à ceux de droite. Cette fois, un effet d'interaction était

bien observé entre l'attitude politique initiale et l'efficacité de la correction sur l'EIC. Supportant l'effet facilitateur d'une correction contre-attitudinale sur l'EIC, la correction réduisait l'adhésion à la FI chez les participants de gauche mais n'avait pas d'effet chez ceux de droite. Le même pattern de résultats a été répliqué dans une seconde expérience (Ecker & Ang, 2019, étude 2).

Par ailleurs, il est aussi important de noter que certaines études ont relevé un effet d'interaction mais allant dans le sens opposé à celui attendu par l'approche motivationnelle, c'est à dire que la correction était *plus* efficace lorsqu'elle était contre-attitudinale. Dans une expérience par exemple (Hameleers et al., 2020), la correction d'une FI qui accusait un réfugié d'être un terroriste était *plus* efficace chez les participants qui avaient des attitudes initiales négatives envers les réfugiés par rapport à ceux qui avaient des attitudes initiales positives (voir aussi Amazeen et al., 2018 ; Young et al., 2018). Face à ce résultat contradictoire parfois observé, une explication est l'occurrence d'un effet plancher lorsque les corrections sont pro-attitudinales. Plus précisément, les participants dont les attitudes initiales s'accordent avec la correction étant ceux qui s'opposent à la FI, ils pourraient ne pas ou n'adhérer que peu à cette dernière. Dans cette situation, l'influence de la correction resterait donc limitée puisqu'elle n'aurait pas d'adhésion sur laquelle être influente. A l'inverse, les participants dont les attitudes initiales s'accordent avec la FI devraient adhérer à la FI. Dans ce contexte, l'influence d'une correction contre-attitudinale pourrait alors être plus grande par rapport à si elle est pro-attitudinale.

En somme, une correction serait généralement moins efficace si elle va à l'encontre des attitudes des individus. Une première condition à cet effet d'interaction est toutefois que la FI corrigée soit suffisamment générale pour remettre en question l'attitude. Une seconde condition est que la FI ait bien une influence chez les individus disposés à s'y opposer. Sans cette dernière condition, un effet d'interaction inverse pourrait s'observer.

4.4 Variables en lien avec la situation

En plus des variables en lien avec les informations, d'autres déterminants peuvent impacter l'EIC. Parmi eux, un aspect important sont les situations dans lesquelles ces informations sont reçues. Par exemple, un temps plus ou moins important peut passer entre la réception de la correction et la sollicitation du modèle. Également, la réception de la FI et de la correction peuvent se faire avec une charge cognitive et un état émotionnel variables.

4.4.1 Effet du temps

Une première variable qui pourrait augmenter l'EIC est le temps passé entre l'intégration de la correction et le moment où le modèle mental est sollicité. De manière générale, une méta-analyse conduite sur les variables modérant l'efficacité d'une correction (Walter & Murphy, 2018) a observé qu'une correction serait moins efficace lorsque l'adhésion à la FI était mesurée une journée ou plus après la correction ($r = 0,13$), par rapport à si elle était évaluée immédiatement ($r = 0,37$) ou après la réalisation d'une courte tâche ($r = 0,48$). Par exemple, dans une recherche (Swire et al., 2017), des participants ont été exposés à une série de FI qui étaient ensuite corrigées. Les participants devaient ensuite réévaluer leur adhésion à ces FI, soit immédiatement après la correction, soit après un délai de 30 minutes ou d'une semaine. Les résultats montraient que l'adhésion aux FI était plus élevée après une semaine par rapport aux autres groupes. De même, plusieurs expériences ont utilisé un paradigme similaire et ont relevé qu'une correction était plus efficace lorsqu'elle était donnée immédiatement après sa réception par rapport à un délai d'une dizaine de minutes (Butler et al., 2023), ou d'une (Ecker, Lewandowsky, et al., 2020) ou plusieurs semaines (Armstrong et al., 1979) (voir aussi Skurnik et al., 2005). Il doit toutefois être noté que cet effet du temps sur l'EIC n'était pas répliqué dans une expérience (Miller et al., 2022).

Une explication possible à l'effet facilitateur du temps sur l'EIC est que le temps passé entre l'encodage d'une information et le moment où elle doit être récupérée réduirait sa capacité

à être récupérée en mémoire. L'effet du temps sur la capacité à récupérer en mémoire un élément est supporté par une expérience (Wixted & Ebbesen, 1991) qui a exposé des participants à une série de questions de connaissances en leur donnant ensuite la bonne réponse. Les mêmes questions étaient à nouveau posées après un délai qui était de six minutes pour un groupe et d'une semaine pour un autre. Supportant l'effet inhibiteur du temps sur la capacité à récupérer une information en mémoire, les participants réalisaient plus d'erreurs si l'évaluation était réalisée une semaine après. Ainsi, dans le cadre de l'EIC, un temps accru entre la sollicitation du modèle mental et l'encodage de la correction pourrait rendre la récupération de cette dernière plus complexe, ce qui réduirait son activation au moment de la compétition d'activation.

Une nuance notable concernant l'effet du temps sur l'EIC est qu'il interagirait avec l'efficacité de la correction. Comme déjà précisé, des manières d'augmenter l'efficacité d'une correction sont, entre autres, de lui donner la forme d'une réfutation ou encore de la répéter. Or, dans une étude (Carnahan et al., 2021), des participants ont évalué leur croyance en une information corrigée immédiatement après sa correction puis après un délai de deux semaines. La correction avait été répétée pour un groupe de participants et ne l'avait pas été pour un autre. Les résultats montraient alors que la réduction de l'efficacité de la correction à la deuxième semaine était moins élevée si la correction avait été répétée, démontrant que la répétition de la correction la rendait moins sensible à l'influence du temps. De même, dans une autre recherche (Swire et al., 2017), des participants ont lu la correction d'une affirmation qui leur avait été présentée puis ont évalué leur adhésion à la FI dans un délai court ou bien d'une ou trois semaines. La correction pouvait être une rétractation ou bien une réfutation. Les résultats montraient un effet d'interaction entre le temps et la formulation de la correction, révélant que le temps avait moins d'impact sur l'EIC lorsqu'une réfutation avait été reçue (voir aussi Ecker, O'Reilly, et al., 2020). En résumé, les corrections efficaces seraient moins sensibles au temps.

Cette nuance offre d'ailleurs une explication à l'absence d'effet du temps sur l'EIC qu'ont observée Miller et al. (2022), les auteurs ayant utilisé une correction qui répétait la FI et lui fournissait une explication alternative.

4.4.2 Charge cognitive

Dans la plupart des études sur l'EIC, les participants disposaient de leurs pleines ressources attentionnelles pour prendre connaissance des informations qui leur étaient présentées. Par exemple, dans l'étude de Johnson et Seifert (1994), les sujets recevaient un texte qu'ils pouvaient lire à leur propre rythme.

Plusieurs expériences suggèrent que les résultats ou la taille d'effet pourraient varier si une charge cognitive est ajoutée lors du traitement des informations. Par exemple, dans une expérience (Sanderson et al., 2023), certains sujets ont écouté des scénarios d'EIC et devaient réaliser des opérations mathématiques en même temps qu'ils recevaient la correction. La FI persistait davantage dans cette condition par rapport à des participants qui n'avaient pas dû réaliser de tâche complémentaire. Une expérience similaire (Ecker, Lewandowsky, Swire, et al., 2011) a demandé à des sujets de mémoriser une série de chiffres en même temps qu'ils prenaient connaissance de la correction d'un texte d'EIC. Les résultats ont montré que l'EIC était plus grand dans cette condition par rapport à une autre ne présentant pas de charge cognitive. Une explication possible à l'effet facilitateur de la charge cognitive au moment de la correction sur l'EIC est d'ordre cognitive. En réduisant la probabilité que la correction soit intégrée au sein du modèle mental, la charge cognitive a pu augmenter l'EIC.

4.4.3 Etat émotionnel

L'influence d'une autre caractéristique situationnelle qui pourrait être interrogée est l'état émotionnel des sujets au moment où ils reçoivent les informations. Sur le plan empirique, il reste toutefois difficile de déterminer clairement l'impact de l'état émotionnel sur l'EIC. D'un côté, une expérience a suggéré un effet inhibiteur d'un état émotionnel négatif sur l'EIC

(Hamby et al., 2020, étude 3). Dans cette expérience, des participants étaient exposés à un texte d'EIC dont l'issue était soit positive (un patient finissait par aller mieux), soit négative (aucun traitement efficace n'était trouvé pour un patient). Une information relative au texte était ensuite rétractée. Les résultats montraient que les inférences en lien avec la FI étaient plus élevées lorsque l'issue du texte était positive, ce qui suggère qu'un état émotionnel positif augmenterait l'EIC. La validité de cette interprétation reste cependant limitée car l'impact des textes utilisés sur l'état émotionnel des participants n'a pas été validé. Par ailleurs, d'autres expériences ont, à l'inverse, suggéré un effet inhibiteur des émotions positives sur l'EIC. Certaines relevaient par exemple que des FI négatives suscitaient un EIC là où des FI positives montraient plutôt un effet de sur-correction (Cobb et al., 2013 ; van Huijstee et al., 2021). Enfin, un autre travail a observé une absence d'effet des émotions sur l'EIC (Ecker, Lewandowsky, & Apai, 2011). Deux groupes de participants ont lu un texte d'EIC relatif au crash d'un avion. Un groupe lisait qu'il était dû à une attaque terroriste et un autre qu'il résultait de la météo. Ces informations étaient ensuite corrigées. La première version était validée comme induisant un état émotionnel plus négatif par rapport à la seconde. Pourtant, l'EIC n'était pas différent entre les deux conditions. Face à ces éléments contradictoires, il reste difficile de déterminer l'influence de l'état émotionnel sur l'EIC. Finalement, comme le relevait une revue de la littérature sur l'EIC (Lewandowsky et al., 2012), la littérature manque encore de travaux autour du rôle des émotions sur l'influence des FI.

Sur le plan théorique, l'hypothèse préférée par McIlhiney, Gignac, Weinborn, et al. (2022) est qu'un état émotionnel négatif inhiberait l'EIC en suscitant du scepticisme, lequel augmenterait l'analyse des informations, qui à son tour réduirait la persistance de la FI. Ce modèle n'a à notre connaissance jamais été évalué mais plusieurs recherches étayaient la validité de certaines de ses propositions. Premièrement, le lien entre le ressenti d'émotions négatives et le scepticisme a été supporté par une étude (Forgas & East, 2008) qui a exposé des participants

à des vidéos dans lesquelles un individu n'aurait pas commis un acte. Dans certaines vidéos, l'acte avait effectivement été commis, tandis que dans d'autres, il ne l'avait pas été. Les participants avaient pour consigne de détecter dans quels films figuraient des mensonges. Avant de réaliser cette tâche, certains participants avaient reçu une induction d'état émotionnel négatif. Supportant l'idée que les émotions négatives augmentaient le scepticisme, les résultats montraient que les participants étaient plus enclins à affirmer qu'ils voyaient des mensonges lorsqu'ils étaient dans un état émotionnel négatif.

Deuxièmement, le lien entre l'état émotionnel négatif et l'analyse des informations a lui aussi été étayé. Dans l'étude de Forgas et East (2008) par exemple, les participants dans un état émotionnel négatif étaient meilleurs pour détecter les tromperies par rapport à ceux qui n'étaient pas dans cet état émotionnel. Également, une expérience (Koch & Forgas, 2012) a montré qu'en règle générale, les phrases qui se lisaient plus facilement avaient plus de probabilités d'être jugées vraies par rapport à celles qui se lisaient plus difficilement ; l'explication étant que les individus se baseraient sur la fluidité dans le traitement des informations pour en estimer leur véracité (Dechêne et al., 2010). Cependant, l'utilisation de cet heuristique semblait disparaître après l'induction d'un état émotionnel négatif. L'interprétation des auteurs était que l'état émotionnel négatif impliquait une analyse accrue des informations, réduisant ainsi le recours aux heuristiques. Des constatations similaires ont été faites dans d'autres travaux observant une réduction dans l'utilisation d'heuristiques (Ruder & Bless, 2003), de stéréotypes (Bodenhausen et al., 1994 ; Lambert et al., 1997) ainsi qu'une augmentation de la latence pour former des jugements (Ruder & Bless, 2003) lorsqu'un état émotionnel négatif était induit.

En montrant un effet de l'état émotionnel négatif sur le scepticisme et l'analyse des informations, ces différentes recherches supportent en partie l'hypothèse de McIlhiney, Gignac, Weinborn, et al. (2022) selon laquelle des émotions négatives pourraient augmenter le

scepticisme, ce qui augmenterait l'analyse des informations, ce qui réduirait l'EIC. L'une des propositions de ce modèle qui reste à déterminer est si une analyse accrue des informations réduit l'EIC, ce qui fera l'objet de la partie suivante.

4.5 Expérience 4. Impact de l'implication sur l'effet d'influence continue

4.5.1 Introduction

L'objectif de la partie 4.5 est d'étudier l'effet de l'implication sur l'EIC. Nous définissons l'implication comme étant une motivation situationnelle à déployer un raisonnement analytique. Selon l'*Elaboration Likelihood Model* (Petty, Barden, et al., 2009 ; Petty & Cacioppo, 1986) ainsi que la théorie des processus duaux (Bohner et al., 1995 ; Craik & Lockhart, 1972 ; de Neys, 2006 ; Evans & Stanovich, 2013 ; Petty, Brinol, et al., 2009 ; Williams et al., 2019), le traitement des informations pourrait être réalisé soit avec un raisonnement intuitif, soit avec un raisonnement analytique (voir cependant Bellini-Leite, 2018). Le premier serait la manière de traiter les informations par défaut, procédant de façon rapide, automatique et intuitive. Le raisonnement analytique consisterait à l'inverse à traiter les informations de façon lente, conscientisée et réfléchie, permettant d'accéder à une autre réponse que la réponse intuitive. Selon la littérature (Petty, Barden, et al., 2009 ; Petty & Cacioppo, 1986), la réunion de deux conditions serait nécessaire pour qu'un raisonnement analytique soit déployé. La première serait d'être en capacité d'utiliser suffisamment de ressources cognitives pour analyser les informations de façon accrue et la seconde d'être suffisamment motivé pour utiliser ces ressources cognitives.

Au sein de la littérature, de nombreux travaux ont montré qu'il était possible de manipuler l'implication de participants lisant un texte en leur précisant par exemple que le texte était lié à des conséquences sûres et immédiates plutôt que lointaines et incertaines (Chaiken & Maheswaran, 1994 ; Petty et al., 1981 ; Petty & Cacioppo, 1981), en écrivant le texte au futur et à la seconde personne du singulier plutôt qu'à la troisième (Burnkrant & Unnava, 1989) et

au passé (Tal-Or & Cohen, 2010) ou encore en leur indiquant qu'il était nécessaire d'être correct dans les réponses aux questions qui suivraient sa lecture (Aarts & Dijksterhuis, 1999 ; Prior et al., 2015 ; Zaccoletti et al., 2023) (voir aussi Bodenhausen et al., 1994 ; Rothman & Schwarz, 1998 ; Wegener & Petty, 1995). Supportant la capacité de ces manipulations à motiver le déploiement d'un raisonnement analytique, ces expériences ont observé qu'elles augmentaient la tendance des participants à savoir rappeler les éléments discutés (Chaiken, 1980 ; Chaiken & Maheswaran, 1994 ; voir cependant Petty et al., 1981) et à comprendre le texte (Zaccoletti et al., 2023), ou encore à être influencé par la qualité argumentative d'un message plutôt que par des heuristiques (Aarts & Dijksterhuis, 1999 ; Bodenhausen et al., 1994 ; Burnkrant & Unnava, 1989 ; Chaiken & Maheswaran, 1994 ; Petty et al., 1981 ; Petty & Cacioppo, 1981 ; Prior et al., 2015 ; Rothman & Schwarz, 1998, 1998 ; Wegener & Petty, 1995).

A partir de l'approche cognitive, il est possible d'émettre l'hypothèse que l'implication pourrait réduire l'EIC. En effet, une analyse accrue des informations devrait réduire le risque que le conflit entre la correction et la FI ne soit pas détecté. De plus, l'implication devrait réduire la probabilité que la correction ne soit pas intégrée faute de ressources cognitives suffisantes. Enfin, si l'analyse accrue des informations se poursuit au moment de la sollicitation du modèle mental, elle pourrait aussi réduire le risque que les individus n'activent pas suffisamment la correction.

D'un point de vue empirique, il existe quelques expériences qui permettent d'aborder l'effet de l'implication sur l'EIC. Une recherche (Ecker et al., 2010) a par exemple montré que des manipulations consistant à rappeler aux participants que certaines informations peuvent manquer de vérifications, ou consistant à réaliser de la psychoéducation sur l'EIC avant de présenter un texte d'EIC, permettaient de réduire cet effet. De la même manière, une autre expérience (Rapp & Kendeou, 2007) a observé une plus grande efficacité de la correction lorsque les participants avaient reçu l'instruction de porter attention aux informations lues. Une

interprétation possible de ces résultats pourrait être que ces manipulations ont stimulé l'implication des participants, ce qui aurait réduit l'EIC. L'influence inhibitrice de l'implication sur l'EIC suggérée par ces expériences reste toutefois hypothétique dans la mesure où l'implication n'a pas été directement contrôlée.

L'objectif principal de l'Expérience 4 a alors été d'examiner si l'implication avait un effet sur l'EIC en veillant à contrôler la manipulation de l'implication. Des participants ont répondu à des questions inférentielles et à une mesure d'adhésion à la croyance pour deux textes, dont l'un contenait une FI qui était subséquemment corrigée tandis que l'autre ne contenait pas de FI. L'EIC était évalué en comparant les réponses entre les deux textes de façon intraindividuelle. Les participants ont aussi été répartis en deux groupes indépendants. Un groupe de participants lisait les textes dans une condition d'implication haute tandis que l'autre les lisait dans une condition d'implication moindre. L'hypothèse émise était que l'EIC devrait être moins élevé lorsque les textes étaient lus dans une condition d'implication haute. Un objectif secondaire de cette étude était de réexaminer la répliquabilité conceptuelle de l'EIC dans le domaine des armées. Pour cette raison, les textes lus décrivaient des situations dans le domaine militaire.

4.5.2 Méthodologie

Le projet de recherche a été validé par le Comité d'éthique en sciences comportementales de l'Université de Lille (référence 2022-648-S111) et a été réalisé en accord avec les standards éthiques de la Déclaration d'Helsinki de 1964. Le consentement éclairé a été obtenu pour tous les participants inclus dans cette étude. Le matériel utilisé est accessible en ligne avec le lien suivant <https://osf.io/bcfej/>.

4.5.2.1 Population

Les conditions d'inclusion étaient d'être majeur et de ne pas être porteur de troubles du spectre de la schizophrénie. La condition de ne pas être porteur de troubles du spectre de la

schizophrénie a été fixée pour éviter l'exacerbation d'une dépersonnalisation que pourrait susciter la lecture d'un texte en condition d'implication haute, demandant d'endosser le rôle d'un personnage.

En cohérence avec des études précédentes sur l'EIC (e.g., Chang et al., 2019 ; Ecker et al., 2011, 2017 ; Johnson & Seifert, 1994 ; Laurent et al., 2023), nous avons décidé de recruter environ 30 participants par groupe indépendant. L'effectif recruté était de 67 participants. Parmi eux, 33 se déclaraient de genre féminin et 2 d'un genre autre. Un total de 44 participants (65,671 %) avait un niveau d'études supérieur au baccalauréat. L'âge moyen était de 33,5 ans ($\sigma = 17,4$). Les participants étaient aléatoirement répartis entre les conditions d'implication haute ($n = 30$) et moindre ($n = 37$). Il n'y avait pas de différence significative dans la répartition des genres ($\chi^2 = 0,762$, $p = 0,683$) ni dans la moyenne d'âge ($U = 456$, $p = 0,211$) entre ces groupes.

4.5.2.2 Matériel

4.5.2.2.1 Textes

Deux textes décrivant un évènement dans un environnement militaire ont été utilisés. Le premier était inspiré du texte utilisé par Johnson et Seifert (1994). Il a néanmoins été légèrement modifié afin de le faire correspondre avec une intervention de pompiers de la Marine nationale sur un bateau. Le deuxième texte décrivait la visite d'un personnage dans une ville ayant subi une attaque. La structure était la même entre les deux textes. Chaque texte se composait d'une série de quinze messages et était précédé d'une phase d'instructions indiquant que la lecture se réalisait sans contrainte de temps et qu'il n'était pas possible de revenir en arrière une fois qu'une page était passée. Chaque texte était aussi précédé de consignes qui variaient en fonction de la condition d'implication du participant.

Les textes pouvaient se décliner selon deux variables. La première était que le texte pouvait comporter une FI qui était ensuite corrigée (version FI) ou bien ne pas en comporter

(version contrôle). Dans la version FI, le sixième message contenait la FI (e.g., des matériaux inflammables auraient contribué à l'incendie du bateau) et le treizième message la corrigeait (e.g., il n'y avait finalement pas de matériaux inflammables). Dans les deux textes en version FI, la source de la correction était la même que la source de la FI. Dans la version contrôle, le sixième message indiquait la vraie information (e.g., il n'y avait pas de matériaux inflammables) et le treizième une information neutre.

La seconde variable manipulait l'implication suscitée par la lecture du texte afin de créer une condition d'implication haute et une condition d'implication moindre. Reprenant des précédents dans la littérature, la manipulation de l'implication était réalisée en modifiant la temporalité (Tal-Or & Cohen, 2010), le pronom personnel (Burnkrant & Unnava, 1989) et les instructions données avant la lecture du texte (Bodenhausen et al., 1994 ; Petty & Cacioppo, 1981 ; Zaccoletti et al., 2023). La manipulation de l'implication est résumée et illustrée dans le Tableau 3.

Tableau 3. Synthèse et illustration de la manipulation de l'implication.

	Implication haute	Implication moindre
Temporalité	Présent	Passé
Pronom personnel	« vous » (singulier)	« il » / « l'officier » / « le capitaine »
Phase d'instructions	<p>Dans ce texte, vous êtes le capitaine d'une section de pompiers de la Marine nationale de votre pays. Il vous est demandé de lire le texte comme si vous étiez un acteur direct de l'évènement qui vous est décrit.</p> <p>Vous êtes invité à lire chaque message avec attention. Des questions sur le texte vous seront posées à la fin de l'expérience. Veuillez lire ce scénario dans le but d'avoir le plus de points possibles.</p>	<p>Dans ce texte, l'un de vos amis s'apprête à vous raconter une histoire vécue par un ancien capitaine d'une section de pompiers de la Marine nationale d'un pays étranger. Il vous est demandé de lire le texte en vous rappelant qu'il s'agit d'une situation fictive. Vous êtes invité à lire le texte comme si vous lisiez un livre. Cette lecture ne suscite aucune obligation de résultats.</p>

4.5.2.2.2 Mesures

Questions inférentielles. Six questions inférentielles étaient posées aux participants (e.g., « *Quelle est la cause la plus probable de l'incendie ?* »). Elles permettaient de calculer un score d'EIC de la même manière que dans l'Expérience 2 (voir partie 3.3.1.2.2).

Adhésion à la croyance. Les participants étaient invités à jauger directement leur adhésion à l'information évoquée dans la FI (e.g., « *Quelle est selon vous la probabilité pour que du gaz et des liquides inflammables aient contribué à l'incendie ?* »). Ils donnaient leur réponse sur une échelle allant de 0 (« *Très faible* ») à 100 (« *Très haute* »). Plus le score était haut, plus il supposait une adhésion au contenu de la FI.

Sentiment d'implication. Afin de confirmer la validité de la manipulation de l'implication, il était demandé aux participants d'indiquer le degré avec lequel le texte qu'ils venaient de lire les impliquait personnellement. Ils se positionnaient sur une échelle allant de 0 (« *Très peu impliqué(e)* ») à 100 (« *Totalement impliqué(e)* »).

4.5.2.3 Procédure

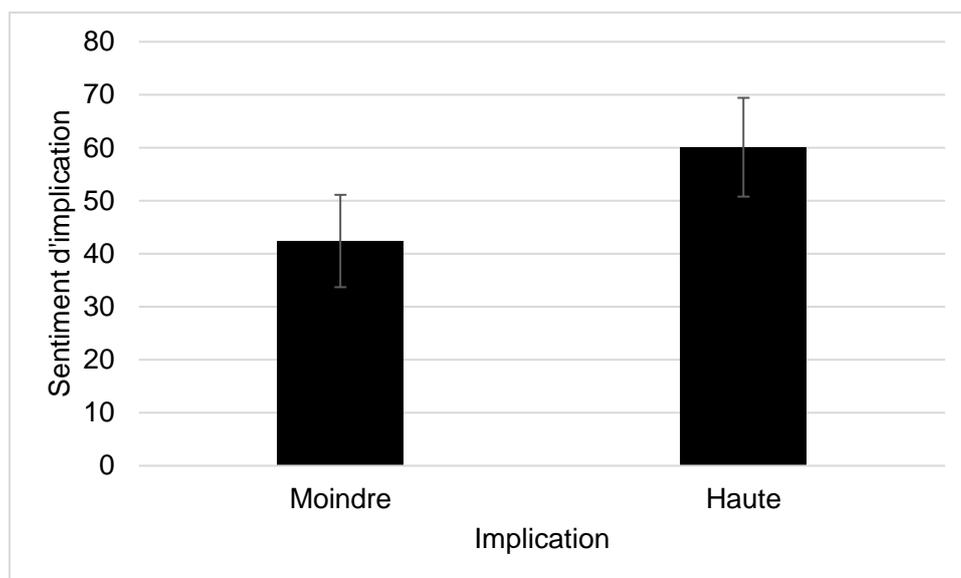
La procédure était informatisée et supportée par le logiciel Inquisit dans sa version web. Les participants étaient recrutés par des annonces diffusées sur différents canaux tels que les réseaux sociaux ou par des affiches. Ils accédaient à l'expérience depuis un lien Internet. Le lancement de la tâche commençait par une lettre d'information qui indiquait que l'étude portait sur le raisonnement envers des textes. Les participants étaient informés qu'ils auraient à lire et répondre à des questions sur deux textes fictifs mettant en scène des événements militaires. S'ils ne s'opposaient pas à prendre part à l'étude, ils signalaient leur genre, leur âge et leur niveau d'études. Au lancement de la passation, le logiciel Inquisit attribuait aléatoirement les participants à l'une des conditions d'implication. Les participants lisaient ensuite les deux textes à leur propre rythme en passant chaque message en appuyant sur la touche « espace » de leur clavier. Ils ne pouvaient pas revenir en arrière dans leur lecture. Chaque message apparaissait sur un écran différent. Le logiciel Inquisit récoltait automatiquement le temps passé sur chaque écran. Tous les participants lisaient un texte en version FI et l'autre en version contrôle. L'ordre des textes et leur version (FI ou contrôle) étaient contrebalancés. Après chaque texte, les questions inférentielles étaient introduites et étaient suivies de la mesure d'adhésion à la croyance puis de la mesure du sentiment d'implication. Chaque mesure apparaissait sur un écran différent. Suite à la dernière mesure, les participants lisaient une page de debriefing qui leur révélait le but de l'expérience. Ils étaient informés de l'utilité qu'auraient leurs données et il leur était laissé le choix d'accepter ou de refuser qu'elles soient conservées à des fins de recherche. Les participants qui refusaient étaient exclus de l'analyse des données.

4.5.3 Résultats

Un test paramétrique était utilisé lorsque les conditions d'applications étaient réunies. Dans le cas contraire, un test non-paramétrique était employé.

4.5.3.1 Vérification de l'implication

Afin d'examiner si la manipulation de l'implication avait été efficace, la moyenne du sentiment d'implication à la lecture des deux textes a été comparée entre les conditions d'implication haute et moindre. Un test de Mann-Whitney a été conduit car la condition de normalité n'était pas respectée ($W = 0,967$, $p = 0,07$). Les résultats montraient une différence significative entre les deux groupes ($U = 358$, $p = 0,013$, corrélation entre rangs bisériés = 0,356). Le sentiment d'implication dans la condition d'implication haute ($\bar{x} = 60,1$, $\sigma = 26,1$) était supérieur à celui ressenti par les participants dans la condition d'implication moindre ($\bar{x} = 42,4$, $\sigma = 27$). Ce résultat est représenté dans la Figure 12.



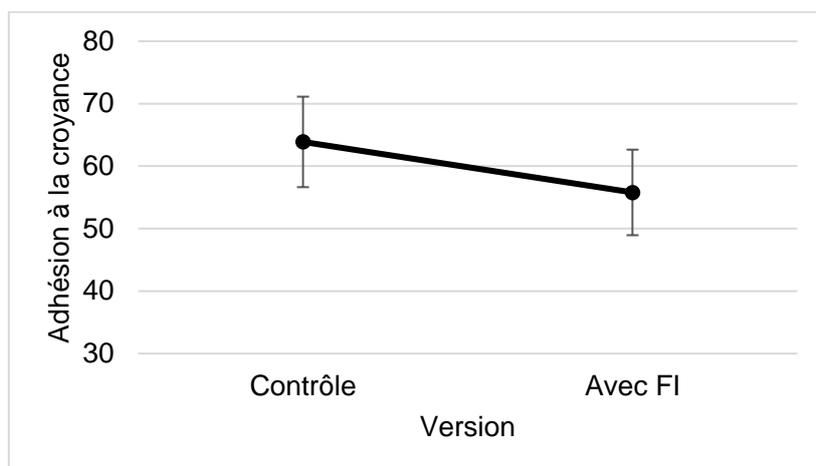
Note. Les barres d'erreur sont des intervalles de confiance, $\alpha = 0,05$.

Figure 12. Score moyen du sentiment d'implication en fonction de la condition d'implication

4.5.3.2 Persistance de la fausse information

Concernant les scores d'EIC, une ANOVA mixte a été réalisée avec la variable FI en mesures répétées (avec FI ; sans FI) et la variable implication en groupes indépendantes (haute ; moindre). L'analyse n'a pas montré de différences de score d'EIC ($F(1,65) = 0,003$, $p = 0,957$) entre les textes en version FI ($\bar{x} = 21,1$, $\sigma = 12,4$) et ceux en version contrôle ($\bar{x} = 21,4$, $\sigma = 10,6$). Il n'y avait pas non plus d'effet d'interaction entre les deux variables ($F(1,65) = 1,333$, $p = 0,252$) ni d'effet principal de l'implication ($F(1,65) = 0,402$, $p = 0,528$).

Concernant la mesure d'adhésion à la croyance, une ANOVA mixte n'a pas été réalisée car la distribution des différences ne respectait pas la condition de normalité ($W = 0,957$, $p = 0,022$). Un test de Wilcoxon a par conséquent été utilisé et a montré une différence significative entre les versions de texte ($W = 674$, $p = 0,022$, corrélation entre rangs bisériés = $-0,331$). L'effet observé ne représentait pas un EIC mais un effet de sur-correction. Les participants adhéraient *moins* au contenu d'une FI lorsque le texte comportait cette FI puis sa correction ($\bar{x} = 55,8$, $\sigma = 28,6$) par rapport à si le texte ne comportait pas de FI ($\bar{x} = 63,9$, $\sigma = 30,3$). Ce résultat est illustré dans la Figure 13.



Note. Les barres d'erreur sont des intervalles de confiance, $\alpha = 0,05$.

Figure 13. Score moyen d'adhésion à la croyance en fonction de la version du texte

A titre exploratoire, il a été examiné si la condition d'implication interagissait avec l'effet de sur-correction observé sur la mesure d'adhésion à la croyance. Pour étudier cet effet d'interaction de façon non-paramétrique, une variable correspondant à la différence des scores d'adhésion à la croyance entre la version FI et la version contrôle a été créée puis a été comparée entre les deux conditions d'implication. Un test de Mann-Whitney a été utilisé en raison du non-respect de la condition de normalité ($W = 0,959$, $p = 0,027$). La comparaison des scores entre la condition d'implication haute ($\bar{x} = -6,47$, $\sigma = 43,1$) et d'implication moindre ($\bar{x} = -9,41$, $\sigma = 32,6$) ne montrait pas de différence significative ($U = 546$, $p = 0,910$).

4.5.4 Discussion

L'objectif principal de cette expérience était d'examiner si l'implication impactait l'EIC. Un objectif secondaire était d'étudier la répliquabilité conceptuelle de l'EIC dans le domaine militaire. Des participants tout-venants ont lu deux textes décrivant des événements dans un contexte militaire. L'un contenait une FI qui était subséquemment corrigée et l'autre ne contenait pas de FI. Un groupe de participants lisait ces textes dans une condition d'implication haute et un autre groupe dans une condition d'implication moindre. Après la

lecture de chaque texte, les participants étaient interrogés sur leur sentiment d'implication personnelle, ce qui a permis de confirmer l'efficacité de la manipulation de l'implication. Ils répondaient également à des questions inférentielles ainsi qu'à une mesure d'adhésion à la croyance. Un EIC n'a pas été observé, empêchant d'étudier l'effet d'interaction attendu. A la place, un effet de sur-correction a été relevé sur la mesure d'adhésion à la croyance.

Une absence d'EIC sur les résultats aux questions inférentielles était inattendue dans la mesure où la méthodologie employée était similaire en de nombreux points au paradigme classique de l'EIC, dont l'Expérience 2 qui montrait cet effet. Rappelons que cette absence d'EIC était également relevée au sein de l'Expérience 1. Une explication possible à l'absence d'EIC observée pourrait être que la FI n'a pas eu d'influence. Cette interprétation semble toutefois peu plausible au regard de l'effet de sur-correction sur l'adhésion à la croyance qui sera ensuite discuté, mais qui suggère déjà que la FI a eu une influence puisque les participants auraient cherché à aller à son encontre. Par conséquent, une explication plausible à l'absence d'EIC sur les résultats aux questions inférentielles est que la FI a eu une influence mais que la correction a ensuite été suffisamment efficace pour permettre, au minimum, un retour des cognitions à l'état de base.

Le caractère efficace de la correction sur les résultats aux questions inférentielles peut être expliqué par plusieurs éléments méthodologiques. Premièrement, une différence majeure entre la présente expérience ainsi que l'Expérience 1 avec les expériences qui observent habituellement cet effet est le caractère militaire des informations. Ce constat suggère que le caractère militaire des informations pourrait réduire la probabilité d'occurrence d'un EIC. Une explication envisageable est que le contexte militaire suscite une implication plus grande par rapport à celle des expériences habituelles sur l'EIC en raison de son association possible à des conséquences importantes. En effet, même pour le groupe d'implication moindre, le sentiment d'implication moyen était de plus de 40 sur 100. Il est possible qu'un texte non-militaire aurait

suscité une implication encore plus faible, ce qui aurait peut-être réduit l'analyse des informations et donc facilité l'EIC. La validité de cette interprétation reste toutefois dépendante de l'effet de l'implication sur l'EIC qui doit encore être démontré.

En plus du caractère militaire des textes, d'autres variables inhérentes à la méthodologie peuvent expliquer l'efficacité de la correction sur les réponses aux questions inférentielles. Par exemple, dans les deux textes, la source de la FI était la même que celle de la correction. Or, dans leur méta-analyse sur des variables modérant l'EIC, Walter et Tukachinsky (2020) relevaient que l'EIC était plus faible voire absent quand la source de la FI était la même que celle de la correction, un effet qu'ils attribuaient à une crédibilité accrue de la correction. D'autres variables pouvant contribuer à l'absence d'EIC sont la pleine disponibilité des ressources attentionnelles, le très court délai entre la réception de la correction et la complétion des mesures ou encore le fait que la correction soit une rétractation qui mentionnait la FI. Ces trois conditions étant identifiées comme des variables qui réduisent l'EIC (Ecker, Lewandowsky, Swire, et al., 2011 ; Kan et al., 2021 ; Swire et al., 2017), il est probable que leur présence dans la méthodologie ait augmenté la probabilité que la correction soit pleinement efficace. Pour autant, il est important de noter que plusieurs expériences ont observé un EIC malgré la présence de ces éléments méthodologiques (e.g., Johnson & Seifert, 1994), ce qui suggère qu'ils ne sont pas suffisants pour expliquer l'absence d'EIC au sein de cette expérience.

Outre l'absence d'EIC, une autre observation importante de cette étude est l'occurrence d'un effet de sur-correction. Lorsque les participants lisaient un texte comportant une FI qui était subséquemment corrigée, ils adhéraient ensuite moins à cette information par rapport à quand ils lisaient un texte sans FI. Selon la littérature autour de l'effet de sur-correction (Wegener & Petty, 1997), l'explication serait que les participants ont détecté l'influence de la

FI sur leur adhésion et qu'ils aient alors surcompensé leurs réponses dans le sens opposé à la FI.

Une nuance notable est que l'effet de sur-correction s'observait sur la mesure d'adhésion à la croyance mais pas sur les scores d'EIC. Une interprétation possible à cette différence est que la mesure d'adhésion à la croyance évalue plus explicitement et en une seule modalité de réponse l'adhésion à la FI par rapport aux scores d'EIC, ce qui a pu faciliter la prise de conscience de l'influence de la FI sur leur jugement. Par exemple, en cherchant à estimer si des matériaux inflammables à l'origine du feu avaient contribué à l'incendie, les participants ont pu penser à une première réponse, puis s'être rappelés que des matériaux inflammables n'avaient pas été retrouvés dans le cagibi, ce qui a pu les mener à réduire leur estimation initiale. A l'inverse, les questions inférentielles telles que « *Quelle est la cause la plus probable de l'incendie ?* » proposaient plusieurs modalités de réponse et leur approche de la FI était plus indirecte. Les diverses modalités de réponse ont pu inciter les participants à analyser de multiples possibilités, réduisant par conséquent leur capacité à utiliser leurs ressources cognitives pour s'auto-corriger. De plus, le lien moins explicite entre les questions inférentielles et la FI a pu être moins propice à ce que les participants détectent que la FI pourrait impacter leur réponse, réduisant donc la probabilité qu'ils cherchent à s'auto-corriger.

En outre, l'absence d'effet d'interaction entre la condition d'implication et l'effet de sur-correction est une observation qui semble aller à l'encontre de ce que prédit la littérature autour de l'effet de sur-correction. En effet, il aurait pu être attendu que les participants dans une condition d'implication haute soient plus enclins à analyser les informations, donc à prendre conscience de l'influence de la FI sur leur jugement et donc à chercher à compenser ce dernier dans le sens opposé de la FI. Une explication possible à cette absence d'effet d'interaction est un effet plafond dans la condition d'implication moindre. Tel qu'indiqué précédemment, la moyenne du sentiment d'implication restait d'environ 40 sur 100 dans cette condition. Il est

donc possible que l'implication soit restée suffisamment haute dans la condition d'implication moindre pour motiver les participants à analyser les informations, prendre conscience de l'influence de la FI et s'auto-corriger.

Une limite importante à l'effet de sur-correction observé est toutefois que la littérature sur ce sujet dans le contexte de la correction des FI reste relativement rare. De plus, il doit être rappelé que cet effet n'était pas attendu dans cette expérience. Ces éléments augmentent le risque de faux positifs et appelle donc à prendre cet effet et son interprétation avec précaution.

De futures études sont nécessaires pour mieux comprendre le rôle de l'implication dans la persistance d'une FI. Concernant l'effet de l'implication sur l'EIC, certaines pourront veiller à utiliser des textes qui ne sont pas militaires, lesquels seront alors peut-être plus propices à l'apparition de cet effet, et pourront ainsi étudier si l'implication modère l'EIC. Des recherches pourront aussi chercher à confirmer l'hypothèse selon laquelle le contexte militaire suscite une implication plus grande, en comparant le sentiment d'implication à la lecture de textes pour lesquels seul le contexte militaire ou non varierait. Enfin, d'autres travaux pourront s'intéresser à l'effet de l'implication sur l'effet de sur-correction en veillant à solliciter une implication haute pour un groupe et une implication très faible pour un autre.

En résumé, cette étude semble observer une correction efficace sur les réponses aux questions inférentielles et un effet de sur-correction sur l'adhésion à la croyance. Cette étude est la deuxième à observer ces effets alors qu'un EIC était attendu (voir Expérience 1). Ce constat suggère que l'importance de l'effet de sur-correction et les corrections efficaces ont probablement été sous-estimées jusqu'alors. Il appelle à de nouveaux travaux autour de l'influence persistante des FI, qui permettraient de mieux comprendre dans quelles conditions et avec quelles populations une correction est efficace, suscite un effet de sur-correction ou bien produit un EIC.

4.6 Expérience 5. Impact de variables dispositionnelles sur l'effet d'influence continue

Une dernière catégorie de variables qui pourraient impacter l'EIC sont relatives aux caractéristiques intrinsèques des personnes qui reçoivent les informations. L'influence de ressources dispositionnelles sur l'EIC est supporté par une recherche (McIlhiney, Gignac, Weinborn, et al., 2022) qui a évalué les inférences des participants sur plusieurs textes d'EIC, permettant de mesurer pour chaque sujet leur sensibilité à cet effet. Quatre semaines plus tard, la procédure a été répétée avec d'autres textes d'EIC, offrant ainsi une deuxième mesure de la sensibilité personnelle des sujets à l'EIC. Alors même que les textes utilisés étaient différents, les résultats ont montré une forte stabilité intra-individuelle, ce qui suggère qu'une partie notable de la variabilité de l'EIC est attribuable à des variables dispositionnelles.

4.6.1 Introduction

A partir de la littérature, il est possible d'identifier plusieurs variables dispositionnelles qui pourraient impacter l'EIC. Nous restreindrons cette partie à l'étude de trois variables dispositionnelles, à savoir la disposition au raisonnement analytique, le névrosisme et la mémoire de travail. L'effet potentiel d'autres variables dispositionnelles sur l'EIC sera évoqué dans la partie 5.2.1.

4.6.1.1 Disposition au raisonnement analytique

Une première variable dispositionnelle qui pourrait impacter l'EIC est la disposition au raisonnement analytique. Un outil fréquemment utilisé pour évaluer cette disposition est le *Cognitive Reflection Test* (CRT, Frederick, 2005). Le CRT est un test composé de trois items, chacun correspondant à un problème mathématique suscitant une réponse intuitive incorrecte. Un exemple d'item est : « Une batte et une balle coûtent 1,10€. La batte coûte 1,00€ de plus que la balle. Combien coûte la balle ? ». La réponse correcte à ce problème est de 0,05 centimes mais l'analyse des réponses montre que la plupart des individus commettent l'erreur de répondre 0,10 centimes, laquelle serait la réponse amenée par un raisonnement intuitif

(Frederick, 2005). Les auteurs suggèrent alors que le nombre de réponses correctes aux items (i.e., le score au CRT) serait une mesure de la disposition au raisonnement analytique : plus le score serait élevé, plus l'individu serait disposé à dépasser sa réponse intuitive en déployant un raisonnement analytique (Toplak et al., 2011). Une expérience (Toplak et al., 2011) a évalué la validité de ce test en faisant passer le CRT à des participants qui ont ensuite dû compléter plusieurs tâches censées susciter des erreurs avec un raisonnement intuitif. Confirmant la validité du test, les résultats ont montré qu'il était un fort prédicteur de la performance à ces différentes tâches (voir aussi Frederick, 2005 ; Trippas et al., 2015).

De façon générale, de nombreux travaux ont mis en évidence une implication de la disposition au raisonnement analytique sur le rapport aux FI. Certaines recherches ont par exemple montré que le raisonnement analytique prédisait une meilleure capacité à différencier les vraies des fausses informations (Bago et al., 2020 ; Barari et al., 2021 ; Bor et al., 2023 ; Brashier et al., 2021 ; Pehlivanoglu et al., 2021 ; R. M. Ross et al., 2021), une moindre propension à partager des FI sur les réseaux sociaux (R. M. Ross et al., 2021) et une plus faible adhésion à des FI politiques (Bronstein et al., 2019).

Dans ce contexte, plusieurs auteurs (Porter & Wood, 2021 ; Xu et al., 2020) ont suggéré que cette influence de la disposition au raisonnement analytique sur le rapport aux FI pourrait se généraliser à l'EIC. D'un point de vue théorique, les arguments sont les mêmes qui étayent l'hypothèse d'un effet inhibiteur de l'implication sur l'EIC. Un traitement accru des informations devrait réduire le risque que le conflit entre la FI et la correction ne soit pas détecté, que la correction ne soit pas intégrée, et que la correction ne soit pas récupérée au moment de la compétition d'activation.

Sur le plan empirique, des travaux précédemment présentés (voir partie 4.5.1) suggèrent qu'une situation stimulant le raisonnement analytique réduirait l'EIC. Cependant, pour rappel, ces évidences restaient indirectes (i.e., elles ne contrôlaient pas l'implication ou le

raisonnement analytique) et situationnelles (Ecker et al., 2010 ; Rapp & Kendeou, 2007). Dans une perspective davantage dispositionnelle, il a été observé que la disposition au raisonnement analytique corrélait négativement avec l'adhésion à des croyances probablement fausses et publiquement corrigées (pour une revue, voir Pennycook et al., 2015), telles que des croyances complotistes (Ståhl & van Prooijen, 2018 ; Swami et al., 2014 ; Vitriol et al., 2022) ou surnaturelles et paranormales (Stagnaro et al., 2019 ; Ståhl & van Prooijen, 2018). Pour autant, soulignons que ces travaux ne manipulaient pas strictement de FI ni de corrections, ce qui ne permet pas de déterminer clairement l'impact de la disposition au raisonnement analytique sur l'EIC.

4.6.1.2 Névrosisme

Selon McIlhiney, Gignac, Weinborn, et al. (2022), une autre variable dispositionnelle qui pourrait impacter l'EIC est le trait de personnalité névrosisme. Le névrosisme désigne la propension à ressentir des émotions négatives quel que soit le niveau de menace objectif posé par la situation (Congard et al., 2012). Les auteurs (McIlhiney, Gignac, Weinborn, et al., 2022) supposaient qu'un individu scorant haut sur une échelle de névrosisme serait davantage prédisposé à être sceptique et donc à analyser les informations.

A notre connaissance, aucun travail n'a cependant directement évalué si le névrosisme prédisait l'EIC. La plupart des évidences relatives au lien entre l'état émotionnel et cet effet sont plutôt d'ordre situationnel (Cobb et al., 2013 ; Ecker, Lewandowsky, & Apai, 2011 ; Hamby et al., 2020 ; van Huijstee et al., 2021). Elles ont été présentées au sein de la partie 4.4.3 et les interprétations qui pouvaient en être tirées étaient difficilement conclusives. Par ailleurs, il doit être souligné que le lien entre une analyse accrue des informations et l'EIC n'a pour l'instant pas été directement démontré, que ce soit de façon situationnelle (voir partie 4.5.1) ou dispositionnelle (voir partie 4.6.1.1). A notre connaissance, le seul travail à avoir abordé le lien entre une disposition à ressentir des émotions négatives et la sensibilité à l'EIC a comparé les

inférences face à un texte d'EIC d'un groupe scorant haut sur des échelles de rumination et de dépression par rapport à un groupe scorant bas sur ces échelles (Chang et al., 2019, étude 1). Les résultats ne montraient pas de différence d'EIC entre les groupes. La capacité à généraliser ce résultat au névrosisme reste toutefois faible car bien que le névrosisme et l'état dépressif soient positivement liés (Saklofske et al., 1995), ils restent des facteurs distincts. En somme, l'hypothèse selon laquelle le névrosisme serait une variable dispositionnelle qui réduirait l'EIC doit encore être évaluée.

4.6.1.3 Mémoire de travail

Une troisième variable dispositionnelle qui pourrait impacter l'EIC est relative aux capacités cognitives des individus. En effet, une expérience (de Keersmaecker et Roets, 2017) a mesuré le niveau d'intelligence de participants qui avaient été exposés à un texte d'EIC. Les résultats ont montré un effet d'interaction entre l'efficacité de la correction et le niveau d'intelligence. Contrairement aux individus aux capacités cognitives les moins élevées, ceux qui avaient les plus hautes capacités cognitives ne montraient pas d'EIC.

Plusieurs auteurs (e.g., Brydges et al., 2018) ont suggéré que l'impact des ressources cognitives sur l'EIC était notamment dû à la Mémoire de Travail (MdT). La MdT est un système, de capacité limitée, qui assure une double fonction de traitement et de stockage temporaire de l'information (Aubry et al., 2018 ; Ehrlich & Delafoy, 1990). Autrement dit, elle est un système qui permettrait de stocker brièvement des informations tout en réalisant une ou plusieurs autres activités mentales (Conway et al., 2005). Par exemple, un individu qui mémorise un numéro de téléphone tout en dialoguant avec un autre activerait sa MdT, puisqu'il stockerait une série de chiffres tout en traitant une conversation.

Dans le cadre de l'EIC, plusieurs éléments théoriques relatifs à l'approche cognitive amènent à émettre l'hypothèse que la MdT pourrait impacter la persistance d'une FI malgré sa correction. Premièrement, il est observé que de meilleures capacités en MdT prédisent une

meilleure compréhension de texte (Daneman & Blennerhassett, 1984 ; Masson & Miller, 1983 ; Zaccoletti et al., 2023), un meilleur raisonnement ou plus généralement une intelligence accrue (de Neys & Verschueren, 2006 ; Furley & Memmert, 2012 ; Jarrold & Towse, 2006). En facilitant le niveau de compréhension, de bonnes capacités en MdT pourraient donc optimiser la clairvoyance des liens entre les différentes informations reçues, ce qui devrait favoriser la détection du conflit entre la correction et la FI. Par ailleurs, corriger une FI après avoir reçu une correction requière d'avoir intégré les deux informations, de remarquer que les deux sont incompatibles, de conclure que la FI ne peut plus être vraie, de déduire qu'elle ne peut plus être utilisée dans la représentation, puis de mettre à jour son modèle mental tout en traitant de nouvelles informations si leur flux est continu. En sollicitant ainsi le stockage d'informations en même temps que leur traitement, il est probable que l'intégration d'une correction dans la mise à jour d'un modèle mental implique la MdT. Par conséquent, de bonnes capacités en MdT pourraient aussi protéger en partie de l'EIC en facilitant l'intégration de la correction (Ecker, Oberauer, et al., 2014).

A notre connaissance, Brydges et al (2018) seraient les premiers à avoir évalué l'effet inhibiteur de la MdT sur l'EIC. Dans leur expérience, des participants ont été exposés à plusieurs textes comportant une FI corrigée ou sans correction, suivis de questions inférentielles et de questions de rappel sur le texte. La MdT des participants était ensuite mesurée. Conformément aux prédictions de l'approche cognitive, les résultats ont montré que le score en MdT prédisait négativement la sensibilité à l'EIC. Plus les capacités en MdT étaient élevées, moins l'EIC était haut. Ce résultat a été répliqué dans les travaux de McIlhiney et al. (2023) ainsi que Jia et al. (2020), qui ont tous deux utilisé des mesures de la MdT différentes.

L'utilisation de différentes mesures de MdT est un élément notable car elle supporte la répliquabilité conceptuelle de l'effet de la MdT sur l'EIC. En l'occurrence, Brydges et al. (2018) ont utilisé trois tâches d'empan complexe. Le principe de ces tâches consiste à présenter

aux sujets des items qu'ils ont pour consigne de stocker en mémoire en même temps qu'il leur est demandé de traiter d'autres informations, comme indiquer si une proposition est correctement écrite, si une opération mathématique est exacte ou encore si des figures sont symétriques. Ces tâches démontrent des qualités psychométriques satisfaisantes et sont souvent utilisées pour évaluer la MdT (Conway et al., 2005 ; Redick et al., 2012). McIlhiney et al. (2023) évaluaient quant à eux l'influence de la MdT sur l'EIC à travers trois autres tâches. Deux d'entre elles étaient des *n-back task*, dans lesquelles les sujets voient des séquences de stimuli (e.g., lettres, formes) présentés individuellement sur des écrans successifs et doivent appuyer sur une touche lorsque le stimulus est le même que celui présenté *n* écrans auparavant. La troisième tâche consistait à présenter de façon successive des mots (e.g., chien, rouge, chat) appartenant à diverses catégories (e.g., animaux, couleurs). A la fin de chaque essai, les participants avaient pour consigne de restituer pour chaque catégorie le dernier mot qui leur avait été présenté. Enfin, Jia et al. (2020) ont utilisé une tâche consistant à rappeler la plus longue série possible de chiffres dans le sens inverse de leur présentation, en plus d'une *n-back task*.

Malgré cette répliquabilité conceptuelle, la validité de l'effet de la MdT sur l'EIC est toutefois remise en question par une étude (Sanderson et al., 2021) qui n'a pas répliqué de lien entre les scores de MdT et l'EIC. Par contre, cette recherche observait que plus les participants répondaient correctement aux questions de rappel, moins ils continuaient de se référer à la FI corrigée. La conclusion des auteurs était que la capacité de mémoire épisodique serait plus importante pour prédire l'EIC que la MdT. Supportant cette explication, les auteurs ont réanalysé les données de Brydges et al. (2018) en incluant dans leurs analyses le score aux questions de rappel qu'avaient eu les participants. Lorsque le score de rappel était ainsi contrôlé, la capacité en MdT ne prédisait plus l'EIC.

Ces évidences contradictoires amènent à questionner la validité de l'effet de la MdT sur l'EIC. Dans ce contexte, l'objectif principal de l'Expérience 5 a alors été d'évaluer le lien

entre la MdT et l'EIC. L'objectif secondaire de cette expérience a été d'évaluer l'effet de la disposition au raisonnement analytique et du névrosisme sur cet effet. Des participants ont été exposés à un texte d'EIC puis ont répondu à des questions inférentielles et à une mesure d'adhésion à la croyance. Ils ont ensuite complété trois mesures de variables dispositionnelles, lesquelles étaient une mesure de la MdT qui n'avait pas encore été utilisée dans la littérature, le CRT et une échelle de névrosisme.

4.6.2 Méthodologie

Le projet de recherche a été validé par le Comité d'éthique en sciences comportementales de l'Université de Lille (référence 2022-648-S111) et a été réalisé en accord avec les standards éthiques de la Déclaration d'Helsinki de 1964. Le consentement éclairé a été obtenu pour tous les participants inclus dans cette étude. Le matériel utilisé est accessible en ligne avec le lien suivant <https://osf.io/bcfej/>.

4.6.2.1 Population

Les conditions d'inclusion étaient d'être majeur et de ne pas être porteur de troubles du spectre de la schizophrénie. Elles étaient aussi de ne pas être diagnostiqué comme étant porteur d'un trouble des apprentissages de type dyslexie afin de ne pas biaiser la mesure de la MdT (Majerus & Poncelet, 2017), ni de type dyscalculique car une telle difficulté dans les opérations mathématiques pourrait biaiser la mesure de la disposition au raisonnement analytique (Patel, 2018).

Nous avons calculé, à l'aide du logiciel G*Power (Faul et al., 2009), la taille d'échantillon pour observer un effet de faible à moyenne taille ($|p| = 0,3$) (en accord avec la littérature sur l'effet potentiel de la MdT sur l'EIC, Brydges et al., 2018) avec une puissance de 0,80. Considérant que l'objectif principal était d'évaluer une covariation entre deux variables (MdT et mesure de l'EIC), une analyse de corrélation a été utilisée pour déterminer la taille d'échantillon. Elle était estimée à 80 participants. L'effectif recruté était de 71 participants.

Parmi eux, 46 déclaraient être de genre féminin et un de genre autre. Un total de 52 participants (75,632 %) avait un niveau d'études supérieur au baccalauréat. L'âge moyen était de 34 ans ($\sigma = 15,5$).

4.6.2.2 Matériel

4.6.2.2.1 Texte

Le texte était une traduction française du texte sur l'incendie de l'entrepôt utilisé par Johnson et Seifert (1994). Il est le même que celui qui avait montré un EIC dans l'Expérience 2 (voir partie 3.3.1.2.1) pour le groupe FI. Pour rappel, le texte se composait de quinze messages. Le sixième message indiquait que des matériaux inflammables stockés négligemment dans un cagibi électrique étaient à l'origine du feu. Le treizième message indiquait que le cagibi qui était supposé contenir des matériaux inflammables était en fait vide lorsque le feu s'est déclenché.

4.6.2.2.2 Mesures

Questions inférentielles. Les questions inférentielles et leur codage étaient les mêmes que celles qui étaient utilisées dans l'Expérience 2. Un score d'EIC était calculé à partir de ces réponses de la même manière que dans l'Expérience 2 (voir partie 3.3.1.2.2).

Adhésion à la croyance. Les participants étaient invités à jauger directement leur adhésion à l'information évoquée dans la FI (e.g., « *Quelle est selon vous la probabilité pour que du gaz et des liquides inflammables aient contribué à l'incendie ?* »). Ils donnaient leur réponse sur une échelle allant de 0 (« *Très faible* ») à 100 (« *Très haute* »). Plus le score était haut, plus il supposait une adhésion au contenu de la FI. La réponse par défaut proposée par le logiciel était 50.

Mémoire de travail. Une tâche d'alpha span employée par Gonthier et al. (2016) a été utilisée afin de mesurer la MdT. Les tâches d'alpha span corrélaient de façon satisfaisante avec des tâches plus classiques d'évaluation de la MdT telles que les tâches d'empan complexe

(Gonthier et al., 2016). La tâche consistait à présenter plusieurs séries de mots francophones, de longueur variable, dont les participants devaient ensuite rappeler chaque première lettre dans l'ordre alphabétique. Par exemple, la bonne réponse à une série telle que « orange », « fauteuil », « dragon » et « langage » serait « DFLO ». Les mots étaient présentés sans majuscule, écrits en noir sur fond blanc, au centre de l'écran. Chaque mot apparaissait pendant deux secondes avant d'être remplacé par un autre. Avant et après le début de chaque série, un cercle rouge apparaissait pendant une seconde au centre de l'écran. Après chaque série, une ligne de cases apparaissait. Le nombre de cases composant la ligne était équivalent au nombre de mots présentés dans la série. Les participants indiquaient leur réponse en inscrivant une lettre par case. Ils devaient obligatoirement commencer par compléter la première case puis la valider avant de pouvoir passer à la suivante. Une fois qu'une case était validée, ils ne pouvaient plus voir leur réponse ni la modifier. Les participants devaient compléter toutes les cases avant de passer à la série suivante. Ils étaient informés qu'ils pouvaient inscrire la lettre « X » s'ils souhaitaient indiquer un vide (aucun mot composant les séries ne commençait par la lettre X). La tâche commençait par une phase d'habitation au cours de laquelle trois séries de mots d'une longueur croissante (i.e., 2, 3, 6) étaient présentées. Après avoir donné leur réponse, un écran indiquait aux participants la bonne réponse. Cette phase d'habitation n'était pas évaluée. La phase suivante était la phase d'évaluation. Elle comprenait huit séries d'une longueur allant de 4 à 8 mots (dans l'ordre : 5, 4, 6, 7, 6, 8, 5, 7). Dans cette phase, les participants n'avaient pas de retour après leurs réponses. Ils obtenaient des points de deux manières. Une première était si la bonne lettre était rappelée dans la bonne case. Ainsi, si la réponse attendue était DFLO mais que la réponse donnée était DLFO ou DXFX, la réponse octroyait deux points. Une deuxième manière d'obtenir des points était si une lettre pouvait être déplacée sur sa position attendue en la décalant sur une case remplie par un X. Ainsi, si la réponse attendue était DFLO, un participant qui répondait XXXD ou XLXX aurait un point, et deux points s'il répondait

XDOX ou LXXO. Le score de MdT correspondait à la somme des points obtenus. Le score minimum était alors de 0 et le score maximum de 48.

Névrosisme. La sous-échelle « névrosisme » de la version française abrégée du test de personnalité NEO-PI-R (Costa & McCrae, 1992, cité dans Poirier, 2012) développée par Sabourin et Lussier (1992, cité dans Poirier, 2012) a été utilisée afin d'évaluer la propension des participants à ressentir des émotions négatives. Cette sous-échelle est composée de douze items évaluant le névrosisme (e.g., « *Je me sens craintif(ve) ou angoissé(e).* ») Les participants indiquaient leur accord sur une échelle de Likert en cinq points allant de 1 (« *Total désaccord* ») à 5 (« *Total accord* »). Les scores étaient codés de sorte à ce qu'un score haut indique un névrosisme plus élevé. Ainsi, les items inversés (e.g., « *Je suis rarement triste ou déprimé(e).* ») étaient recodés en soustrayant 6 par le score brut. Le névrosisme était évalué en calculant la somme des scores aux items.

Raisonnement analytique. Une mesure inspirée du CRT (Frederick, 2005) a été utilisée afin d'évaluer la disposition au raisonnement analytique des participants (Pennycook et al., 2016). En raison de la médiatisation des items initiaux du test (e.g., Gentside, 2018 ; Zylberberg, 2018) qui pourrait dégrader la validité de la mesure (Patel, 2018), une modification des termes des problèmes a été effectuée. Par exemple, le problème de la batte et de la balle a été modifié par « *Le coût d'un plat et d'un dessert est de 60 euros. Le plat coûte 50 euros de plus que le dessert. Combien coûte le dessert ?* ». Une telle modification des termes a déjà été réalisée dans la littérature (Patel, 2018). Les trois items apparaissaient sur un même écran. Chaque item était suivi d'un encadré dans lequel le participant pouvait inscrire sa réponse. Le scoring était réalisé en attribuant un point pour chaque bonne réponse (Pennycook et al., 2016). Ainsi, la note minimale était de 0 et le score maximal de 3.

4.6.2.3 Procédure

La procédure était informatisée et supportée par le logiciel Inquisit dans sa version web. Les participants étaient recrutés par des annonces diffusées sur les réseaux sociaux leur permettant d'accéder à l'étude depuis un lien Internet. Une lettre d'information indiquait que l'étude portait sur le raisonnement envers des textes. Les participants étaient informés qu'ils auraient à lire un texte fictif se composant d'une série de messages, répondre à des questions sur la lecture puis compléter une tâche d'évaluation cognitive, un questionnaire et un test. S'ils consentaient à prendre part à l'étude, ils signalaient leur genre, leur âge et leur niveau d'études. Chaque participant lisait ensuite le texte. Après quoi, ils répondaient aux questions inférentielles puis à la mesure d'adhésion à la croyance, lesquelles étaient suivies, dans un ordre aléatoire, des mesures de MdT, de névrosisme et de raisonnement analytique. Suite à la dernière mesure, les participants lisaient une page de debriefing leur révélant le but de l'expérience. Ils étaient informés de l'utilité qu'auraient leurs données et il leur était laissé le choix d'accepter ou de refuser qu'elles soient conservées à des fins de recherche. Les participants qui refusaient étaient exclus de l'analyse des données.

4.6.3 Résultats

4.6.3.1 Mesures de l'effet d'influence continue

Le score d'EIC moyen était de 20,8 ($\sigma = 8,87$). Ce score était significativement différent de 0 ($t(70) = 19,7$, $p < 0,001$). Par ailleurs, le score d'adhésion à la croyance moyen était de 63,3 ($\sigma = 26,1$). La condition de normalité pour la distribution de ce score n'étant pas respectée ($W = 0,936$, $p = 0,001$), un test de Wilcoxon a été utilisé pour déterminer s'il était significativement différent de la réponse par défaut. L'analyse montrait que le score d'adhésion à la croyance était significativement plus haut ($W = 1641$, $p < 0,001$) par rapport à ce que proposait la réponse par défaut.

4.6.3.2 Effet des variables dispositionnelles sur l'effet d'influence continue

Les coefficients de corrélation linéaire entre les variables dispositionnelles, le score d'EIC et la mesure d'adhésion à la croyance ont été calculés. Ils sont rapportés dans le Tableau 4.

Tableau 4. Corrélations entre les mesures

	1.	2.	3.	4.	5.
1. Score EIC	-				
2. Adhésion à la croyance	0,455**	-			
3. Mémoire de travail	-0,252*	-0,018	-		
4. Névrosisme	-0,128	-0,071	0,173	-	
5. CRT	-0,075	-0,188	0,283*	-0,054	-

Note. * $p < 0,05$, ** $p < 0,01$.

La seule variable dispositionnelle qui montrait une relation significative avec l'une des mesures de persistance de la FI était la MdT. Le score de MdT corrélait négativement avec le score d'EIC ($p = 0,034$) mais pas avec la mesure d'adhésion à la croyance ($p = 0,881$). Une analyse de régression linéaire a été conduite avec le score d'EIC en variable dépendante et celui de MdT en variable indépendante ($B\grave{e}ta = -0,210$, $p = 0,034$, $R^2 = 0,063$). Le nuage de points et la droite des moindres carrés décrivant la relation entre ces deux variables sont présentés dans la Figure 14.

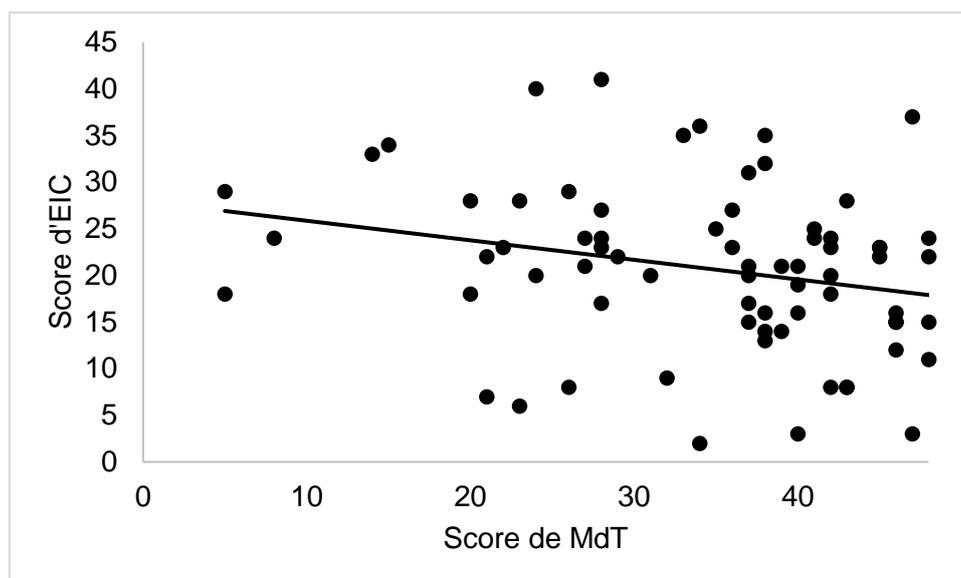


Figure 14. Nuage de points et droite des moindres carrés de la relation entre le score en MdT et le score d'EIC

4.6.4 Discussion

Dans le cadre de cette expérience, nous avons émis les hypothèses que les scores en MdT, au CRT et à l'échelle de névrosisme devraient être négativement associés à l'EIC. Des participants ont été exposés à un texte dont la capacité à susciter un EIC a précédemment été démontrée (voir Expérience 2). Ils ont ensuite répondu à des questions inférentielles mesurant un score d'EIC ainsi qu'à une mesure d'adhésion à la croyance. Après quoi, les participants ont complété une tâche de MdT, une échelle de névrosisme et le CRT. Le score en MdT était négativement lié au score d'EIC mais pas à celui d'adhésion à la croyance. Une relation linéaire n'a pas été observée entre les mesures de la persistance de la FI et le névrosisme ainsi que le CRT.

L'objectif principal de cette expérience était d'étudier le lien entre la MdT et l'EIC. Allant à l'encontre d'une précédente étude n'ayant pas relevé d'effet de la MdT sur l'EIC (Sanderson et al., 2021), les analyses ont montré un lien négatif entre la MdT et le score d'EIC : plus le score de MdT augmentait, moins les participants avaient tendance à attribuer des points de plausibilité aux propositions en lien avec la FI. En accord avec l'approche cognitive, une interprétation théorique de cet effet est que le processus d'intégration de la correction sollicite

la MdT. Par conséquent, plus les individus auraient une MdT élevée, plus ils seraient efficaces pour mettre à jour leur modèle mental, et moins ils seraient donc à risque de montrer un EIC. Par ailleurs, une explication complémentaire pourrait être que l'amélioration de la compréhension associée à la MdT (Daneman & Blennerhassett, 1984 ; Masson & Miller, 1983 ; Zaccoletti et al., 2023) a facilité la détection du conflit entre la FI et la correction. En tout cas, ce résultat s'accorde avec la littérature qui avait déjà observé une relation négative entre la MdT et l'EIC (Brydges et al., 2018 ; Jia et al., 2020 ; McIlhiney et al., 2023). Il contribue de plus à la répliquabilité conceptuelle de cette relation en ayant employé une tâche de MdT qui n'avait pas encore été utilisée dans la littérature sur l'EIC.

Une observation notable concernant la relation entre la MdT et l'EIC est que cette relation était faible. En fait, cette relation faible mais significative est en accord avec d'autres travaux autour de l'EIC. Par exemple, l'étude de Brydges et al. (2018) relevait qu'environ 7 % de la variance de leur mesure d'EIC était prédite par leur modèle de MdT. Par ailleurs, il peut être noté que le paradigme employé au sein de la présente expérience était peu propice à ce que la MdT exerce une grande influence sur la persistance de la FI. En effet, les participants disposaient de leurs pleines ressources attentionnelles pour lire le texte et pour répondre aux questions. De plus, le texte ne comportait qu'une unique FI qui était ensuite corrigée. Dans d'autres conditions, des individus pourraient recevoir plusieurs FI avant d'apprendre en même temps que plusieurs d'entre elles étaient fausses. Par exemple, Drummond et al. (2020) ont exposé leurs participants à de multiples titres d'articles de journaux fallacieux avant de leur révéler, en un unique message, que toutes ces informations étaient fausses. Dans une telle situation impliquant de traiter simultanément plusieurs mises à jour de modèles mentaux, il est probable que l'effet de la MdT sur la sensibilité à l'EIC soit plus prononcé.

Une question est de déterminer pourquoi l'effet de la MdT a été observé sur les scores d'EIC et pas sur la mesure d'adhésion à la croyance. Une explication possible est que la mesure

d'adhésion à la croyance est une question explicite sur l'adhésion à la FI (e.g., « *Quelle est selon vous la probabilité pour que du gaz et des liquides inflammables aient contribué à l'incendie ?* »). En raison de son caractère explicite, il est possible que cette mesure constitue une nouvelle occasion pour les participants de prendre conscience que l'information initiale a été corrigée et donc d'intégrer la correction au sein de leur modèle mental. Par conséquent, cette mesure semble moins propice pour détecter un EIC qui serait dû à un défaut d'intégration de la correction au moment de sa réception. À l'inverse, les questions inférentielles sont des mesures plus implicites de l'EIC. Elles n'interrogent pas directement l'adhésion à la FI et proposent plusieurs modalités de réponse. Elles sont donc moins propices à susciter une nouvelle prise de conscience de la correction de la FI au moment d'y répondre, ce qui expliquerait pourquoi l'effet de la MdT sur l'EIC a été observé pour les questions inférentielles mais pas pour l'adhésion à la croyance.

Des hypothèses secondaires de cette étude étaient que le névrosisme et le CRT devraient prédire la persistance d'une FI en réduisant le risque que le conflit entre la FI et la correction ne soit pas détecté ou pour que la correction ne soit pas intégrée ni récupérée faute de ressources cognitives. Une première explication à l'absence d'effet de ces variables dispositionnelles pourrait être que ces postulats théoriques sont erronés. Ils reposent toutefois sur de nombreux travaux (voir parties 4.6.1.1 et 4.6.1.2), ce qui rend cette interprétation discutable. Une autre explication est que le paradigme utilisé n'a pas été fortement propice à l'émergence d'une influence du névrosisme et de la disposition au raisonnement analytique sur l'EIC. Premièrement, la FI mentionnant la présence de matériaux inflammables dans le cagibi était explicite. De plus, la correction était une rétractation qui mentionnait la FI. L'utilisation d'une FI implicite (Rich & Zaragoza, 2015) et d'une correction ne mentionnant pas la FI auraient pu augmenter la difficulté pour détecter le conflit entre la FI et la correction, ce qui aurait pu donner une influence plus grande à l'analyse des informations sur l'EIC.

Deuxièmement, les mesures de persistance de la FI étaient données immédiatement après la lecture du texte, soit peu de temps après la lecture de la correction. La conséquence est que la plupart des participants, qu'ils soient ou non disposés à analyser les informations, ont pu aisément récupérer la correction. Dans un autre paradigme incluant un délai plus long entre la réception de la correction et la sollicitation du modèle mental, l'EIC pourrait davantage dépendre de la récupération de la correction et donc des efforts déployés pour la récupérer, ce qui pourrait donner un rôle plus important à l'analyse des informations dans cet effet.

Cette interprétation basée sur des éléments méthodologiques reste toutefois spéculative. Elle devrait être confirmée par de futurs travaux qui pourront par exemple manipuler le délai entre la correction et la sollicitation du modèle pour étudier si les influences du névrosisme et de la disposition au raisonnement analytique sur l'EIC sont plus importantes lorsque le délai est long. De futurs travaux pourront aussi manipuler le caractère explicite de la FI et la mention de la FI dans la correction afin de déterminer si ces variables interagissent avec l'effet du névrosisme et du raisonnement analytique sur l'EIC. De façon plus générale, d'autres études pourront s'intéresser à l'effet d'autres variables dispositionnelles sur l'EIC, telles que le besoin de clôture cognitive (D. M. Webster & Kruglanski, 1994), le besoin en cognition (Cacioppo & Petty, 1982) ou encore la disposition à avoir une pensée active et ouverte (Svedholm-Häkkinen & Lindeman, 2018). Ces variables seront discutées dans la partie 5.2.1.

Une limite de cette expérience est qu'elle n'incluait pas de groupe contrôle dans lequel la FI n'aurait pas été présentée, ce qui ne permet pas de confirmer directement la présence d'un EIC. Plusieurs arguments réduisent toutefois l'importance de cette limite. Le premier argument est relatif à la similarité de la méthode et des résultats entre cette expérience et l'Expérience 2. Dans l'Expérience 2 qui avait montré un EIC, le texte en version FI, les questions inférentielles et la procédure étaient identiques en tous points par rapport à ce qui a été utilisé dans cette expérience. De plus, le score d'EIC qui avait été obtenu dans l'Expérience 2 pour le groupe en

version FI était similaire ($\bar{x} = 21,667$; $\sigma = 9,525$) à celui obtenu ici ($\bar{x} = 20,8$; $\sigma = 8,87$). Supportant leur similarité, un test de Student comparant ces deux groupes ne montrait pas de différence significative ($t(90) = -0,398$, $p = 0,691$). Enfin, un test de Student a été conduit entre les scores d'EIC des participants de cette expérience et ceux du groupe contrôle dans l'Expérience 2 ($\bar{x} = 15,8$; $\sigma = 7,56$). Les analyses montraient alors une différence significative entre les deux groupes ($t(92) = -2,43$, $p = 0,017$), ce qui supporte également l'occurrence d'un EIC. Un autre argument à l'occurrence d'un EIC est que les scores variaient en cohérence avec ce que prédit la littérature, comme l'atteste la prédictivité des scores en MdT sur le score d'EIC. En l'absence d'EIC, un tel résultat aurait eu peu de probabilités d'être observé. Enfin, nous avons mené des analyses démontrant que la moyenne des scores d'EIC était significativement supérieure à 0 et supérieure à 50 pour celle d'adhésion à la croyance. Bien que ces résultats ne permettent pas d'affirmer que les participants faisaient plus de références à la FI par rapport à s'ils ne l'avaient pas reçue, ils permettent d'attester que les participants faisaient plus de références à la FI par rapport à s'ils ne l'envisageaient pas du tout probable, ce qui s'accorde avec ce qu'engendrerait un EIC.

En conclusion, le présent travail a mis en évidence une relation négative entre la MdT et la persistance d'une FI corrigée. La poursuite des études autour des variables dispositionnelles qui impacteraient l'EIC est importante pour plusieurs raisons. Sur le plan théorique, l'identification de ces déterminants permet de mieux comprendre comment varie l'EIC et ce qui le détermine. Sur le plan pratique, leur identification peut permettre d'optimiser le recrutement pour des postes dans lesquels l'environnement informationnel est changeant, tels que des analystes du renseignement, des opérateurs de l'aéronautique ou encore des cadres. En bref, chaque individu serait plus ou moins sensible à l'EIC, et l'identification de ceux qui le sont davantage offrirait de prévenir les répercussions négatives qu'il peut avoir.

Chapitre 5. Conclusion

Au sein d'un premier chapitre introductif, nous avons cherché à démontrer en quoi et à quel point les FI seraient une problématique contemporaine majeure. Après avoir proposé une définition des FI applicable au présent travail, leur importance au sein de la société a été étudiée à travers leur diffusion et l'adhésion qu'elles suscitent, ainsi qu'en examinant les multiples domaines qu'elles impactent. Il a ensuite été exposé qu'une stratégie fréquemment employée contre l'influence des FI consiste à les corriger. Cette introduction apparaissait nécessaire pour situer l'importance d'étudier l'EIC aujourd'hui. En remettant en question l'efficacité des corrections, cet effet interroge la pertinence d'une stratégie communément employée pour faire face à une problématique sociétale majeure.

Un second chapitre a cherché à définir ce qu'est l'EIC. Nous nous sommes pour cela appuyés sur le paradigme classique de l'EIC (e.g., Johnson & Seifert, 1994). De plus, à partir d'une analyse de la littérature conduite dans le cadre de cette thèse, nous avons proposé une catégorisation inédite des différentes explications théoriques de cet effet en trois approches : cognitive, motivationnelle et rationnelle. Enfin, nous avons aussi présenté des effets voisins de l'EIC en détaillant d'autres issues potentielles à la suite de la correction d'une FI, à savoir un effet boomerang, un effet de sur-corrrection, ou encore un retour ou maintien de la ligne de base.

Dans le troisième chapitre, nous avons étudié la validité de l'EIC et avons pour cela suivi deux pistes complémentaires. Une première consistait à faire le point sur la répliquabilité conceptuelle de cet effet. Observant finalement un effet de sur-corrrection, les résultats d'une première étude publiée dans *Applied Cognitive Psychology* (Laurent et al., 2024) ne sont pas parvenus à répliquer un EIC dans un contexte militaire. Ce constat s'insère toutefois au sein d'une littérature riche et diverse autour de l'EIC qui permet malgré cela d'en assurer la répliquabilité conceptuelle générale. Une piste complémentaire pour étudier la validité de cet effet a consisté à en étudier la répliquabilité proche, un point qui n'était encore que peu développé dans la littérature. Une répliquabilité proche du paradigme classique de l'EIC (Johnson & Seifert,

1994) a alors été conduite avec une population francophone, dans une étude publiée dans *Acta Psychologica*. Le succès de cette réplique a permis, à son tour, d'étayer la validité de cet effet.

Au sein du quatrième chapitre, nous avons cherché à relever des déterminants importants de l'EIC. A partir de l'analyse de la littérature, nous avons choisi de les diviser en quatre catégories : ceux en lien avec les FI, ceux en lien avec la correction, ceux relatifs à la situation dans laquelle les informations sont reçues, et ceux liés aux personnes qui reçoivent les informations. Nous avons cherché à enrichir la littérature en menant trois études complémentaires. L'une d'elles a souhaité explorer si une FI sous forme de *deepfake* pouvait engendrer un EIC et si cet effet était plus grand par rapport à une FI sous forme de texte, mais le paradigme employé n'a pas permis de susciter un EIC. Une autre recherche a étudié si l'implication situationnelle dans le traitement des informations pouvait réduire l'EIC, mais un effet de sur-correction a été observé à la place de cet effet. Un troisième travail a examiné l'effet de plusieurs variables dispositionnelles sur la sensibilité à l'EIC. Les résultats ont permis de supporter la littérature en répliquant le lien entre la MdT et cet effet. Au sein de cette dernière partie, nous souhaitons poursuivre l'étude de l'EIC en synthétisant les apports théoriques et pratiques qu'espère apporter le présent travail, ainsi qu'en esquisant les perspectives pour ceux à venir.

5.1 Apports

5.1.1 Apports théoriques

De premiers apports théoriques touchent au fonctionnement de l'EIC. Au sein de la partie 2.2, nous avons proposé une conceptualisation inédite des explications théoriques les plus fréquentes autour de l'EIC en proposant de les regrouper en trois approches : cognitive, motivationnelle et rationnelle. Cette conceptualisation constitue à notre connaissance la première tentative de synthèse théorique de cet effet. La validité de l'approche cognitive a de

plus été étayée par l'Expérience 4 qui a montré une relation négative entre les capacités en MdT et le score d'EIC, un résultat qui supporte notamment l'hypothèse qu'un processus d'intégration de la correction au sein du modèle mental interviendrait dans l'EIC.

Une autre contribution théorique que permet ce travail concerne la validité de l'EIC. L'analyse de la littérature menée dans la partie 3.1 couplée au succès d'une réplique proche de cet effet (voir partie 3.3) étayent fortement l'idée que les cognitions peuvent rester impactées dans le sens d'une FI malgré sa correction. De plus, l'Expérience 2 observant un EIC avec une population francophone permet de supporter la validité de cet effet en montrant son caractère transculturel.

L'observation de la validité de l'EIC s'insère dans un contexte scientifique où l'effet boomerang est quant à lui fortement remis en question (voir partie 2.3.1.3). Alors que plusieurs scientifiques s'étaient inquiétés de la possibilité qu'une correction puisse polariser les cognitions d'une personne dans le sens de l'information corrigée (Cook & Lewandowsky, 2012 ; Nyhan et al., 2014), par effet de familiarité ou bien de vision du monde, le consensus actuel semble maintenant que ce phénomène serait suffisamment rare pour ne pas devoir inquiéter les individus qui voudraient corriger des FI (Ecker et al., 2023 ; Swire-Thompson et al., 2020 ; Wood & Porter, 2016). Ainsi, l'état de l'art actuel est que les échecs de réplique de l'effet boomerang (Ecker et al., 2023 ; voir Swire-Thompson et al., 2020) se multiplient en même temps que les succès de réplique de l'EIC (e.g., McIlhiney et al., 2023 ; Rich et al., 2022 ; Sanderson et al., 2022). Ensemble, la forte répliquabilité de l'EIC couplée à l'importante remise en question de l'effet boomerang pourrait laisser penser qu'après la correction d'une FI, l'EIC serait le phénomène le plus souvent observé.

Plusieurs observations au sein de ce travail appellent toutefois à nuancer l'idée d'une primauté de l'EIC. Bien qu'une de nos expériences ait démontré cet effet (Expérience 2) et qu'une autre ait fourni des évidences allant dans son sens (Expérience 5), il est important de

noter que les trois autres ne l'ont pas observé. Ces différentes expériences reprenaient pourtant en grande partie les éléments méthodologiques principaux du paradigme classique de l'EIC, à savoir l'utilisation d'une succession de messages suivie de questions inférentielles.

Une explication possible à l'absence d'EIC dans ces recherches est un retour des cognitions à la ligne de base, c'est-à-dire que la correction aurait été suffisamment efficace pour évincer l'influence de la FI. Une autre explication à l'absence d'EIC pourrait être un maintien de la ligne de base, c'est-à-dire que la FI n'aurait pas eu d'influence. Cette seconde explication ne peut pas être strictement exclue car les expériences réalisées n'utilisaient pas de groupe contrôle avec FI et sans correction qui aurait permis de confirmer l'influence de la FI. Toutefois, l'hypothèse d'une absence d'influence de la FI reste peu probable dans les Expériences 1 et 4 qui ont montré un effet de sur-correction, lequel suggère que la FI aurait eu une influence puisque les individus auraient cherché à aller à son encontre. Quoiqu'il en soit, l'absence de réplication de l'EIC dans plusieurs expériences suggère que l'EIC serait davantage ténu et sensible à des variations du paradigme informationnel que ne le laisse entendre l'état de l'art. En somme, une première conclusion de ce travail serait que l'EIC est un effet valide mais qu'il reste important de ne pas surestimer la probabilité de son occurrence à la suite de la correction d'une FI.

Par ailleurs, l'observation d'un effet de sur-correction au sein de ce travail est un autre apport théorique majeur. Cet effet est encore peu exploré dans la littérature relative aux FI (voir cependant Cobb et al., 2013 ; Sommers & Kassin, 2001 ; van Huijstee et al., 2021) et encore très peu de recherches font l'hypothèse de son occurrence. Son observation dans au moins deux recherches différentes suggère que sa probabilité d'apparition à la suite de la correction d'une FI a potentiellement été jusqu'alors sous-estimée. Par conséquent, une deuxième conclusion est que l'effet de sur-correction à la suite de la correction d'une FI est peut-être plus valide et plus

probable que ce que laissait paraître la littérature, bien que cette proposition reste à confirmer par de futurs travaux.

5.1.2 *Apports pratiques*

D'un point de vue pratique, la validité de l'EIC a des implications majeures dans un contexte sociétal où la problématique des FI est notable et que de nombreux acteurs s'attachent justement à les corriger (voir Chapitre 1). L'examen de la validité de cet effet confirme que les acteurs qui luttent contre les FI doivent activement prendre en compte qu'une simple correction peut ne pas être suffisante pour évincer leur influence.

La nécessité de prendre en compte l'EIC s'applique aux organisations civiles (Graves & Cherubini, 2016) ou politiques (e.g., NATO, 2020) qui corrigent des FI. Sur le plan scientifique, il est pertinent de relever que l'EIC pose aussi une problématique éthique pour les recherches qui manipulent des FI. Certains travaux ont enrichi la littérature en présentant des FI à des participants au sujet des vaccins (de Saint Laurent et al., 2022 ; Featherstone & Zhang, 2020 ; Jolley & Douglas, 2014 ; Loomba et al., 2021 ; Porter & Wood, 2021), du changement climatique (Bolsen & Druckman, 2018 ; Porter & Wood, 2021 ; van der Linden, 2015) ou de la peine de mort (Cohen et al., 2000). L'EIC suggère que ce type de méthodologie pourrait continuer d'influencer les cognitions des participants, même avec l'utilisation d'un debriefing révélant le caractère factice des informations (Greitemeyer, 2014 ; Misra, 1992).

L'analyse de la littérature menée dans ce travail permet de proposer, aux acteurs de l'approche corrective et aux chercheurs manipulant des FI, certaines pistes pour réduire l'EIC. Premièrement, la partie 4 a permis de mettre en avant plusieurs déterminants qui réduiraient directement cet effet, tels que le fait de répéter la correction (Carnahan et al., 2021 ; Ecker, Lewandowsky, Swire, et al., 2011 ; Sanderson et al., 2023 ; Vraga & Bode, 2017a), de veiller à la crédibilité de la correction (Butler et al., 2023 ; Ecker & Antonio, 2021 ; Guillory & Geraci, 2013 ; O'Rear & Radvansky, 2020 ; Vraga & Bode, 2017a) et à la disponibilité des ressources

cognitives au moment de sa réception (Ecker, Lewandowsky, Swire, et al., 2011 ; Sanderson et al., 2023), ainsi que l'utilisation d'une réfutation reprenant les faits ou ajoutant une explication alternative (Autry & Duarte, 2021 ; Ecker et al., 2010 ; Ecker, Lewandowsky, & Apai, 2011 ; Jin et al., 2022 ; O'Rear & Radvansky, 2020 ; Rich et al., 2022 ; Swire et al., 2017). Des expériences (Ecker et al., 2010 ; Kan et al., 2021 ; Paynter et al., 2019) ont d'ailleurs observé que la combinaison de ces variables semblait optimiser l'efficacité de la correction. Par exemple, Ecker et al. (2010) ont relevé que le fait de fournir une explication alternative à la FI tout en incitant les participants à porter leur attention sur l'EIC était plus efficace pour réduire cet effet par rapport à si une seule de ces manipulations était employée. Les acteurs qui luttent contre les FI de même que les chercheurs qui en manipulent pourraient se saisir de l'identification de ces déterminants proposée dans ce travail pour optimiser leurs corrections ou leurs briefings (Greene et al., 2023). Ainsi, un premier apport pratique de ce travail est d'avoir fait la synthèse de plusieurs éléments permettant de rendre une correction plus efficace. A notre connaissance, une telle synthèse était absente de la littérature.

Par ailleurs, ce travail a proposé une synthèse des recherches autour de l'approche rationnelle, lesquelles démontrent l'importance de la crédibilité de la correction pour réduire l'EIC. Cette synthèse permet de proposer qu'une piste indirecte pour réduire l'EIC serait d'accroître la crédibilité des sources de correction, telles que les médias d'information (Graves & Cherubini, 2016) et le monde scientifique (Nyhan & Reifler, 2010 ; Vlasceanu & Coman, 2022 ; Vraga & Bode, 2020). Cette perspective a une signification contemporaine particulière dans la mesure où une confiance moindre ou déclinante envers ces sources de correction est justement constatée par plusieurs enquêtes. En effet, entre 32 % (CSA, 2022 ; IPSOS, 2024) et 68 % (Boy, 2021) voire 72 % (CEVIPOF, 2024) de la population déclarerait ne pas faire confiance aux médias d'information. Concernant la confiance envers la science, en 2024, environ 20 % d'un échantillon de la population française affirmait ne pas avoir confiance en

cette institution (CEVIPOF, 2024), soit une augmentation de sept points par rapport à 2011 (Boy, 2014 ; voir aussi Kraus et al., 2023). Dans ce contexte et en cohérence avec l'approche rationnelle, un autre apport pratique de ce travail est la recommandation qu'une manière de réduire la persistance des FI au sein de la société serait de travailler sur la confiance recueillie par les sources de correction.

Enfin, en ayant dressé une synthèse de l'approche cognitive (voir partie 2.2.1) et supporté sa validité (voir partie 4.6.4), ce travail permet d'émettre l'hypothèse qu'une piste pour réduire l'EIC toucherait à la communication autour de cet effet au sein de la population. En effet, l'approche cognitive suggère qu'une analyse accrue des informations pourrait favoriser l'encodage et l'intégration de la correction. Or, il est possible qu'être éduqué sur l'EIC rende les individus plus vigilants lorsqu'ils sont dans une situation où une information est corrigée, ce qui suppose que la psychoéducation sur l'EIC pourrait réduire cet effet. L'intérêt de communiquer autour de l'EIC pour le prévenir est supporté par une étude (Ecker et al., 2010) dans laquelle un groupe de participants a été exposé à un texte d'EIC après avoir reçu un message de psychoéducation sur cet effet. Ces participants montraient moins de références à la FI corrigée par rapport à un autre groupe n'ayant pas reçu cette psychoéducation. Ainsi, un autre apport pratique que permet l'analyse de la littérature conduite dans ce travail est la recommandation de communiquer sur l'EIC. L'application de cette dernière recommandation devra toutefois dépendre des résultats de recherches futures.

5.2 Perspectives

5.2.1 Recherches futures

A partir de l'analyse de la littérature conduite dans ce travail et des recommandations pratiques élaborées, plusieurs pistes peuvent être proposées pour de futures recherches. Premièrement, concernant la communication autour de l'EIC, il est important que les bénéfices de cette recommandation soit validée par d'autres travaux avant d'être exécutée. Pour cause,

malgré ses intérêts théoriques, cette communication pourrait aussi exposer au développement de l'effet de sur-correction. En effet, le fait de rendre les individus conscients qu'une FI pourrait continuer d'influencer leurs cognitions malgré sa correction pourraient les amener à s'auto-corriger de manière excessive s'ils surestiment l'importance de l'EIC. Des recherches sont donc nécessaires pour déterminer les bonnes manières de communiquer sur l'EIC sans produire un effet de sur-correction.

De façon plus générale, d'autres travaux sont nécessaires autour des déterminants de l'EIC. Malgré l'identification de beaucoup d'entre eux (voir Chapitre 4), il reste encore difficile de prédire précisément quand cet effet survient plutôt qu'un effet de sur-correction ou une correction efficace. L'identification de nouveaux déterminants de cet effet et de ce qui le réduit sera utile aux acteurs luttant contre l'influence des FI. Comme précisé précédemment, des déterminants de l'EIC qui restent à étudier sont, entre autres, l'implication, l'état émotionnel ou encore la valence de la FI. Également, comme le rappelaient Lewandowsky et al. (2012), la connaissance des variables dispositionnelles qui prédisent l'EIC reste moindre et de futures recherches sont nécessaires pour continuer d'étudier cette perspective.

L'analyse de la littérature nous a permis de relever plusieurs variables dispositionnelles dont l'étude pourrait être intéressante. Une première est le besoin de clôture cognitive (« *need for closure* »), lequel fait référence au « désir pour une réponse donnée, quelle qu'elle soit, par rapport à la confusion et à l'ambiguïté » (D. M. Webster & Kruglanski, 1994). Il est évaluable par la *Need for Closure Scale* (NFCS, Webster & Kruglanski, 1994), qui se compose de plusieurs sous-dimensions telles que la préférence pour l'ordre et la structure ou encore l'inconfort généré par l'ambiguïté. En accord avec l'approche motivationnelle, une hypothèse possible serait qu'en déstabilisant la structure du modèle mental, la correction d'une FI susciterait davantage d'inconfort chez les individus hauts en besoin de clôture et que ces individus seraient donc plus sensibles à l'EIC. Des éléments indirects qui supportent cette

hypothèse sont que la NFCS (D. M. Webster & Kruglanski, 1994) corrèle positivement avec une échelle de dogmatisme (Webster & Kruglanski, 1994, étude 2), ou encore que les individus hauts en besoin de clôture ont moins tendance à changer leur évaluation initiale d'une personne malgré de nouvelles informations (Webster & Kruglanski, 1994, étude 4, 6).

Une autre variable dispositionnelle qui pourrait intéresser les futurs travaux sur l'EIC est le besoin en cognition (« *need for cognition* »), qui désigne la « tendance à s'engager et prendre plaisir à penser dans une grande diversité de domaines » (Cacioppo & Petty, 1982 ; Frederick, 2005 ; Petty, Brinol, et al., 2009). Une échelle qui évalue cette variable est la *Need for Cognition scale* (NC, Cacioppo & Petty, 1982), qui utilise plusieurs items auto-rapportés tels que « *Je préfère les tâches complexes aux tâches simples.* » (Cacioppo & Petty, 1982). En abordant la tendance à déployer des ressources cognitives, cette mesure apparaît proche du CRT. Elle doit toutefois en être distinguée car elle est une mesure auto-rapportée alors que le CRT est une mesure de performance. De plus, la NC évalue la propension à s'engager dans des pensées, mais cette motivation ne garantit pas leur caractère analytique (Petty, Brinol, et al., 2009). Dans le cadre de l'EIC, de futurs travaux pourraient étudier si la NC prédit cet effet. Un individu qui apprécie s'engager dans des activités cognitives serait peut-être plus enclin à consacrer des ressources cognitives aux informations reçues et donc à pouvoir détecter le conflit entre la FI et la correction, intégrer la correction, ou récupérer cette dernière lors de la sollicitation du modèle mental.

Une troisième variable dispositionnelle qui pourrait intéresser la recherche sur l'EIC est la disposition à avoir une pensée active et ouverte (« *Actively Open-minded Thinking, AOT* »), qui se caractérise par la disposition à chercher et à s'engager de manière réfléchie avec de nouvelles informations même si elles peuvent être menaçantes (J. Carpenter et al., 2018). Cette variable est évaluée par l'*AOT-Scale* (Svedholm-Häkkinen & Lindeman, 2018), une échelle composée de plusieurs sous-dimensions dont le dogmatisme, la résistance aux faits, le

libéralisme et le refus de juger les individus pour leurs opinions (Svedholm-Häkkinen & Lindeman, 2018). Dans le cadre de l'EIC et en accord avec l'approche motivationnelle, des personnalités moins dogmatiques ou plus ouvertes face aux évidences contradictoires pourraient être plus enclines à modifier leur modèle mental même quand une correction va à l'encontre de leurs attitudes initiales.

Outre l'étude de variables dispositionnelles, un autre point auquel pourraient veiller les futurs travaux sur l'EIC serait d'intégrer des variables comportementales (Courchesne et al., 2021 ; de Saint Laurent et al., 2022 ; Murphy et al., 2023). La plupart des mesures dans les études sur l'EIC sont des questions inférentielles ou des mesures d'adhésion à la croyance. La réponse à ces questions ne semble toutefois pas totalement dépendre des mêmes processus que des variables (d'intention) comportementales. Par exemple, Nyhan et al. (2014) ont observé que la correction du lien entre un vaccin et l'autisme réduisait l'adhésion à cette croyance chez des participants, mais augmentait en même temps leur hésitation vaccinale. De la même manière dans l'Expérience 1, un effet de sur-correction était relevé sur les prises de décision mais pas sur la réponse aux questions inférentielles. L'intégration de variables comportementales ou d'intention comportementale dans de futures recherches permettrait donc d'enrichir la littérature sur l'EIC.

Une autre recommandation pour les futures recherches serait d'accroître le caractère écologique des études sur l'EIC. Mosleh et al. (2022) suggéraient notamment de développer l'utilisation des réseaux sociaux comme terrain d'expérience directe, une piste déjà explorée par quelques travaux dans la littérature relative aux FI. Par exemple, une recherche a observé que l'envoi d'un message privé demandant à un internaute d'évaluer la véracité d'un article amenait à une augmentation de la véracité des tweets subséquentement diffusés (Pennycook et al., 2021). Également, Bowles et al. (2020) ont diffusé des messages d'information sur le COVID-19 dans des groupes WhatsApp durant la période de confinement, ce qui aurait eu

l'effet d'augmenter le respect de la distanciation sociale chez les récepteurs. Dans le cadre de l'EIC, de futures expériences pourraient s'inspirer de ces recherches pour créer des paradigmes plus écologiques.

Par ailleurs, comme l'ont relevé Courchesne et al. (2021), la plupart des études portant sur les FI ont lieu sur des participants issus de démocraties occidentales. Or, comme précisé précédemment (voir partie 3.3), la culture pourrait impacter l'EIC (Xu et al., 2020). La conduite de futurs travaux qui compareraient l'EIC entre différentes cultures apporterait des intérêts théoriques si elle identifie des différences. Ces travaux pourront aussi être utiles pour connaître la pertinence plus ou moins grande de l'approche corrective en fonction de l'endroit où elle est appliquée. En 2021, des FI circulant sur les réseaux sociaux accusaient des soldats français de piller de l'or au Mali et d'y fournir des armes aux groupes terroristes, ce qui aurait participé aux mouvements de blocage des convois militaires par la population locale (Chahed, 2021 ; Ianni, 2021 ; Mandjo, 2021). Connaître la sensibilité de cette population à l'EIC aurait permis aux communicants militaires de mieux déterminer à quel point l'approche corrective était une réponse adaptée ou si d'autres stratégies devaient être envisagées.

Une autre recommandation portée par Ecker et Antonio (2021) et en cohérence avec l'analyse de la littérature menée dans ce travail serait de porter davantage d'attention à l'approche rationnelle pour expliquer l'EIC. Au sein de la littérature, l'approche rationnelle est rarement reprise pour rendre compte de cet effet. Pourtant, plusieurs éléments montrent qu'un défaut de crédibilité de la correction pourrait souvent expliquer une part des EIC. Premièrement, trois recherches ont directement interrogé la croyance en la correction de participants qui venaient de lire un texte d'EIC. Les résultats relevaient que la moitié des participants déclaraient ne pas croire en la correction (O'Rear & Radvansky, 2020) et qu'autour d'un tiers des participants estimaient que la correction qui leur était donnée dissimulait quelque chose (Guillory & Geraci, 2010) ou qu'elle était un mensonge (Guillory & Geraci, 2013). Ainsi, ces

travaux montrent que lorsque la crédibilité d'une correction est interrogée, celle-ci semble souvent dégradée, ce qui suggère qu'un défaut de crédibilité de la correction pourrait être fréquent dans les EIC observés dans la littérature. Un autre argument qui supporte l'importance de l'approche rationnelle est l'hypothèse selon laquelle la correction aurait, par défaut, un handicap de crédibilité par rapport à la FI (Ecker & Antonio, 2021). La raison avancée par les auteurs est qu'une (fausse) information ne serait pas forcément remise en question lorsqu'elle est reçue. À l'inverse, le correcteur souffrirait d'un handicap de crédibilité car, par essence, il contredit quelque chose qui était probablement considéré comme vrai. Le correcteur susciterait alors de la méfiance là où le désinformateur n'en rencontrait pas (Connor Desai et al., 2020). Supportant l'hypothèse d'un handicap de crédibilité, une expérience a demandé à des participants d'évaluer la crédibilité de chaque message dans plusieurs textes d'EIC et a observé que la crédibilité de la correction était en moyenne moins haute par rapport à celle de la FI (Ecker & Antonio, 2021). Ainsi, l'hypothèse du handicap de crédibilité de la correction (Connor Desai et al., 2020 ; Ecker & Antonio, 2021) en plus des recherches qui montrent que la correction manque fréquemment de crédibilité (Ecker & Antonio, 2021 ; Guillory & Geraci, 2010, 2013 ; O'Rear & Radvansky, 2020) appellent à porter davantage de considération à l'approche rationnelle au sein de la littérature sur l'EIC (Ecker & Antonio, 2021).

Pour finir, une autre piste pour de futurs travaux serait d'étudier le lien encore inexploré entre la correction des FI et la modification de croyances dysfonctionnelles, que nous définissons comme étant des croyances rigides dont l'expression est à la source d'une dégradation de l'adaptation à l'environnement. Par exemple dans le cadre de l'insomnie, des illustrations de telles croyances seraient « *je devrais dormir aussi bien que mon compagnon* » ou encore « *si je me couche après minuit, je ne dors pas* » (Hartley et al., 2023). Plusieurs recherches ont observé qu'une réduction de l'adhésion à des croyances dysfonctionnelles était associée à une amélioration de l'état clinique (Grant et al., 2012 ; Hartley et al., 2023). Il serait

intéressant de déterminer à quel point les connaissances autour de la correction des FI acquises dans la littérature autour de l'EIC seraient transposables à la modification des croyances dysfonctionnelles. De futurs travaux pourraient par exemple évaluer si une croyance dysfonctionnelle en clinique est mieux modifiée lorsque les contre-arguments du thérapeute sont répétés, agrémentés d'explications alternatives ou encore lorsqu'ils sont fournis au début de l'entretien afin d'optimiser la disponibilité des ressources cognitives du patient.

5.2.2 Limite commune aux expériences

Les limites des différentes expériences ont déjà été discutées dans leurs parties respectives, c'est pourquoi seule une limite commune aux expériences est présentée ici. Elle se réfère à la représentativité du niveau d'études des échantillons recrutés. Au sein de la population française, la proportion d'individus diplômés du supérieur serait d'environ 50 % (INSEE, 2020). Ce taux était au-dessus de 65 % dans la plupart des expériences (61 % dans l'Expérience 2). Il est possible que ce défaut de représentativité ait influencé les résultats sur l'EIC. Pour cause, le niveau d'études serait positivement lié à l'intelligence et ceci de façon modérée voire forte (Deary et al., 2007). Or, l'intelligence serait pour sa part associée à une réduction de l'EIC (De keersmaecker & Roets, 2017). De même, le niveau d'études serait associé à un plus haut besoin en cognition (Liu & Nesbit, 2024), une variable qui pourrait également être associée à une réduction de l'EIC (voir partie 5.2.1). Par conséquent, il est possible que la taille d'effet de l'EIC et que son occurrence au sein des expériences soient sous-estimées par rapport à la population typique.

5.2.3 Limites et alternatives à l'approche corrective

Ce travail s'est principalement porté sur l'EIC. Cet effet constitue une limite notable et valide de l'approche corrective en montrant que l'influence d'une FI peut persister dans le sens de la FI malgré sa correction. Par ailleurs, plusieurs résultats de ce travail ont précisé qu'une autre limite de l'approche corrective pouvait être l'effet de sur-correction. Cet effet est

une autre limite de l'approche corrective en montrant qu'une correction peut aussi amener les individus à être biaisés en opposition excessive à la FI corrigée. Enfin, il peut être souligné ici que l'effet boomerang ne constituerait qu'une limite négligeable à l'approche corrective en raison de la faible probabilité de son occurrence que relève la littérature.

Outre l'EIC et l'effet de sur-correction, d'autres limites notables de l'approche corrective existent. Une première est que ce sont les individus qui auraient le plus besoin de recevoir une correction qui y seraient les moins exposés sur Internet. Supportant cette proposition, une recherche (Melki et al., 2021) a observé que parmi les individus qui avaient partagé des informations sur le COVID-19, ceux qui adhéraient à des fausses croyances sur ce sujet avaient aussi moins tendance à s'engager dans des démarches permettant de vérifier le caractère véridique de ces informations (e.g., vérifier la source originale, vérifier l'expertise de la source). De même, une autre étude a relevé que les individus qui aimaient les pages complotistes sur Facebook interagissaient très peu avec les contenus de vérification des faits (Zollo et al., 2017). Par ailleurs, une autre limite à l'approche corrective est qu'elle prend du temps. Un rapport publié en 2016 identifiait que des équipes de vérifications des faits produiraient en moyenne moins de sept contenus par semaine (Graves & Cherubini, 2016). Bien que la rapidité de leur production semble s'être accrue en 2024 (e.g., <https://factuel.afp.com>), elle reste largement inférieure aux quelques minutes nécessaires pour créer et partager une FI en ligne. Ainsi, cet état d'asymétrie temporelle couplé à la portée restreinte de l'approche corrective, l'EIC et l'effet de sur-correction, obligent à envisager des alternatives (Alemanno, 2018) pour réduire l'influence des FI au sein de la société.

Une première alternative à l'approche corrective pour lutter contre les FI pourrait reposer sur des solutions technologiques. Par exemple, Kumar et Geethakumari (2014) ont proposé le développement d'un algorithme capable de détecter des tweets à risque d'être des FI. Dans la même idée, d'autres auteurs (Molina et al., 2021) ont suggéré de développer des

algorithmes capables de déterminer si une information diffusée sur un site était une fausse nouvelle à partir de plusieurs indicateurs tels que l'analyse sémantique de l'information, l'examen de l'URL ou encore du contenu associé à l'onglet « *about us* ». L'ambition de ces auteurs est que ces algorithmes étiquettent automatiquement les informations à risque d'être fausses, ce qui permettrait un gain de temps notable par rapport à l'approche corrective classique consistant à apporter manuellement une correction. Cette étiquetage automatique des FI resterait toutefois limité par l'insuffisance des corrections (Treen et al., 2020), d'autant plus sous cette forme d'étiquette (Ecker, O'Reilly, et al., 2020). Il risquerait aussi de rester limité par la crédibilité moindre qu'une source de correction pleinement technologique pourrait avoir.

Une autre alternative possible est la stratégie d'inoculation, qui consiste à agir en amont de l'exposition à une FI afin de construire une résistance cognitive à cette FI chez les individus (Basol et al., 2021 ; Roozenbeek, van der Linden, et al., 2020 ; Traberger et al., 2022). Plusieurs matériaux disponibles en libre accès ont été développés sur ce principe, tels que les jeux *Go viral!* (www.goviralgame.com), *Bad News* (www.getbadnews.com) ou encore des vidéos (Roozenbeek et al., 2022). Avec ces outils, les participants étaient invités à découvrir ou à pratiquer directement les techniques d'influence souvent observées dans la diffusion de fausses nouvelles (e.g., citer des faux experts, recourir à un langage émotionnel, créer des TC). Les expériences manipulant ces outils ont observé qu'ils étaient efficaces pour améliorer la discrimination entre les vraies et les fausses informations et pour réduire l'adhésion à des FI ainsi que la tendance à les partager (Basol et al., 2020, 2021 ; Roozenbeek et al., 2022 ; Roozenbeek, van der Linden, et al., 2020). L'avantage de cette action d'inoculation est qu'elle apparaît généralisable à la plupart des fausses nouvelles et ne requièrent que peu de temps une fois le matériel créé, ce qui offre une alternative intéressante à l'approche corrective. Une proposition de Roozenbeek et al. (2022) était par exemple de diffuser des vidéos d'inoculation dans les campagnes publicitaires de YouTube. Une limite à ces outils est toutefois qu'ils gardent

une utilisation très générale, ce qui ne permet pas d'envisager l'application de cette approche à une FI spécifique dont l'influence devrait être évitée.

Une autre alternative à l'approche correctrice résiderait dans l'éducation à l'esprit critique (Bronner et al., 2022 ; Swami et al., 2014). L'esprit critique est entendu comme « la capacité à ajuster son niveau de confiance de façon appropriée selon l'évaluation de la qualité des preuves à l'appui et de la fiabilité des sources » (Pasquinelli & Bronner, 2021). Le postulat est que l'éducation à l'esprit critique permettrait d'induire un scepticisme sain chez les individus, les rendant plus enclins à l'analyse des informations, donc à la détection de celles qui sont fausses et donc à la réduction de l'adhésion qu'elles récoltent. En ce sens, il a été observé que l'esprit critique était associé à une moindre adhésion à des TC (Lantian et al., 2021) et que le simple fait de poser une question sur la véracité d'une TC réduisait la propension des participants à l'endosser (Einstein & Glick, 2015). De même, il a été relevé que la tendance à vérifier les informations avant de les partager sur les réseaux sociaux prédisait négativement le niveau d'adhésion à des FI (Melki et al., 2021) ou encore que l'exposition à des vidéos d'aide à l'identification des FI (e.g., vérifier les URL, la source de l'information) contribuait à un meilleur discernement pour distinguer les vraies vidéos par rapport à des fausses et augmentait la qualité épistémique des informations subséquentement partagées (Bor et al., 2023). Afin d'améliorer l'éducation à l'esprit critique, Pasquinelli et Bronner (2021) ont notamment proposé de l'intégrer de façon transversale dans différentes matières de l'enseignement primaire, secondaire et supérieur. Cette approche reste toutefois limitée par le temps qu'elle requière pour la formation des enseignants et pour être dispensée au sein des cours. Elle est aussi plus difficilement applicable à des individus sortis du système scolaire. Enfin, ici aussi, le caractère général de cette approche ne permet pas d'envisager son application à la lutte contre l'influence d'une FI spécifique.

Une dernière approche complémentaire envisageable pourrait concerner la régulation des informations diffusées. Premièrement, cette régulation peut consister à interrompre la diffusion d'informations fausses. Cette piste a été investie par des plateformes telles que TikTok (Brewster et al., 2022) ou Instagram. Par exemple, en 2018, le hashtag #911wasdonebythejews aurait été supprimé par Instagram (Reichstadt & Igounet, 2018). Cette piste est aussi investie dans le cadre législatif. Ce dernier énonce, entre autres, que la diffusion d'une FI peut être punie et interrompue si elle trouble ou est susceptible de troubler la paix publique, notamment si elle s'applique au domaine militaire (Article 27 de la loi du 29 juillet 1881), si elle porte atteinte aux intérêts de la Nation pour servir un acteur extérieur (Article 411-10 du Code Pénal), ou encore si elle survient dans des périodes proches des élections présidentielles (Loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information). L'interruption de la diffusion des FI est une piste qui apparaît pertinente pour réduire la diffusion des FI au sein de la société. Elle reste toutefois limitée par le fait de ne pas agir sur l'adhésion des individus qui y ont été exposés. Elle pose de plus des questions éthiques relatives à la liberté d'expression (Badouard, 2018 ; Treen et al., 2020). Outre l'interruption de la diffusion des FI, une régulation plus nuancée pourrait consister à réduire sa visibilité. Par exemple, Facebook fait appel à ses utilisateurs et à des organismes indépendants pour détecter les FI. Une fois qu'un contenu est ainsi identifié comme faux, la plateforme le fait apparaître plus bas dans le fil d'actualités (*À propos de la vérification des informations sur Facebook*, 2022). Cette approche reste toutefois limitée par le fait qu'il est nécessaire que la FI soit identifiée, ce qui requière du temps et implique qu'elle a probablement déjà été largement diffusée avant d'être rétrogradée.

En conclusion, l'approche corrective a plusieurs limites qui sont sa portée restreinte, le temps qu'elle requière, l'EIC et l'effet de sur-correction. Face aux limites de l'approche corrective, il semble nécessaire de recourir à des alternatives telles que des solutions technologiques, l'inoculation, l'éducation à l'esprit critique ou encore la régulation des

informations. Toutefois, malgré leurs intérêts, ces alternatives restent elles aussi limitées. Ainsi, la meilleure stratégie pour lutter contre les FI semble être de développer ces approches de façon complémentaire, dont l'approche corrective. Certes, l'EIC est un effet valide et il est donc nécessaire de reconnaître qu'une correction peut parfois s'avérer insuffisante pour évincer l'influence d'une FI. Toutefois, l'EIC est également un effet variable, dépendant de multiples déterminants relatifs à la FI, la correction, la situation et la personne qui reçoit les informations. Cette variabilité, appuyée par les résultats de cette thèse, offre un message nuancé mais encourageant aux acteurs de l'approche corrective : sous certaines conditions, une correction peut être pleinement efficace. Depuis 2021, la dynamique des organisations de vérifications des faits, qui avait progressé dans les années 2010 (Stencel & Griffin, 2018), semble stagner (Stencel et al., 2023). La relance de leur expansion apparaît cruciale dans un contexte où la diffusion des FI, l'adhésion qu'elles suscitent et les conséquences qu'elles génèrent se poursuivent.

Références

- À propos de la vérification des informations sur Facebook.* (2022). Pages d'aide Meta Business. <https://fr-fr.facebook.com/business/help/2593586717571940>
- Aarts, H., & Dijksterhuis, A. (1999). How often did I do it? Experienced ease of retrieval and frequency estimates of past behavior. *Acta Psychologica*, 103(1), 77-89. [https://doi.org/10.1016/S0001-6918\(99\)00035-9](https://doi.org/10.1016/S0001-6918(99)00035-9)
- Adriani, R. (2019). The Evolution of Fake News and the Abuse of Emerging Technologies. *European Journal of Social Sciences*, 2(1), 18-24. <https://doi.org/10.26417/ejss-2019.v2i1-53>
- Ahmed, S. (2021). Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism. *New Media & Society*, 146144482110191. <https://doi.org/10.1177/14614448211019198>
- Alemanno, A. (2018). How to Counter Fake News? A Taxonomy of Anti-fake News Approaches. *European Journal of Risk Regulation*, 9(1), 1-5. <https://doi.org/10.1017/err.2018.12>
- Allard-Huver, F. (2017). Fake news. In *Publictionnaire : Dictionnaire encyclopédique et critique des publics*. <http://publictionnaire.huma-num.fr/wp-content/uploads/2017/11/fake-news.pdf>
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. <https://doi.org/10.1257/jep.31.2.211>
- Alley, S., Jennings, C., Persaud, N., Horsley, M., Plotnikoff, R., & Vandelanotte, C. (2014). Do Personally Tailored Videos in a Web-Based Physical Activity Intervention Lead to Higher Attention and Recall? – An Eye-Tracking Study. *Frontiers in Public Health*, 2(13), 1-7. <https://doi.org/10.3389/fpubh.2014.00013>
- Allgaier, J. (2019). Science and Environmental Communication on YouTube: Strategically Distorted Communications in Online Videos on Climate Change and Climate Engineering. *Frontiers in Communication*, 4(36), 1-15. <https://doi.org/10.3389/fcomm.2019.00036>
- Amadiou, J.-F. (2002). *Le poids des apparences : Beauté, amour et gloire*. Odile Jacob.
- Amazeen, M. (2015). Revisiting the Epistemology of Fact-Checking. *Critical Review*, 27(1), 1-22. <https://doi.org/10.1080/08913811.2014.993890>
- Amazeen, M., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting Political and Consumer Misperceptions: The Effectiveness and Effects of Rating Scale Versus Contextual Correction Formats. *Journalism & Mass Communication Quarterly*, 95(1), 28-48. <https://doi.org/10.1177/1077699016678186>
- Anderson, L. E. (2010). *Operation Fortitude: The Allied D-Day deception campaign and media use*. [University of Wisconsin-Eau Claire]. https://minds.wisconsin.edu/bitstream/handle/1793/44601/Anderson_Luke.pdf?sequence=1

- Armstrong, G. M., Gurol, M. N., & Russ, F. A. (1979). Detecting and Correcting Deceptive Advertising. *Journal of Consumer Research*, 6(3), 237-246.
<https://doi.org/10.1086/208765>
- Aubry, A., Gonthier, C., Hainselin, M., & Bourdin, B. (2018). *Measuring the development of working memory capacity: Example of an adaptive complex span task* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/uxbt7>
- Autry, K. S., & Duarte, S. E. (2021). Correcting the unknown: Negated corrections may increase belief in misinformation. *Applied Cognitive Psychology*, 35(4), 960-975.
<https://doi.org/10.1002/acp.3823>
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556-559. <https://doi.org/DOI:10.1126/science.1736359>
- Baddeley, A. (2010). Working memory. *Current Biology*, 20(4), 136-140.
<https://doi.org/10.1016/j.cub.2009.12.014>
- Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, 63(1), 1-29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Badouard, R. (2018). Les « “fake news” » menacent-elles le débat public ? *Les Cahiers Français*, 406, 36-47.
- Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, 149(8), 1608-1613. <https://doi.org/10.1037/xge0000729>
- Ballew, M. T., Leiserowitz, A., Roser-Renouf, C., Rosenthal, S. A., Kotcher, J. E., Marlon, J. R., Lyon, E., Goldberg, M. H., & Maibach, E. W. (2019). Climate Change in the American Mind: Data, Tools, and Trends. *Environment: Science and Policy for Sustainable Development*, 61(3), 4-18.
<https://doi.org/10.1080/00139157.2019.1589300>
- Barari, S., Lucas, C., & Munger, K. (2021). *Political Deepfake Videos Misinform the Public, But No More than Other Fake Media*. OSF Preprint.
- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W., & van der Linden, S. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, 8(1), 20539517211013868. <https://doi.org/10.1177/20539517211013868>
- Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good News about Bad News: Gamified Inoculation Boosts Confidence and Cognitive Immunity Against Fake News. *Journal of Cognition*, 3(1), 1-9. <https://doi.org/10.5334/joc.91>
- Basso, B., & Kanellos, I. (2019). Infox, réalités construites et post-vérité : Diachronies d'un phénomène si moderne. In *The construction of reality in the post-truth era* (Alma Mater). <https://shorturl.at/3lpWo>
- Bellais, R. (1998). Les enjeux de la maîtrise de l'information dans la Défense. *Réseaux. Communication - Technologie - Société*, 16(91), 121-133.
<https://doi.org/10.3406/reso.1998.3210>
- Bellini-Leite, S. C. (2018). Dual Process Theory: Systems, Types, Minds, Modes, Kinds or Metaphors? A Critical Review. *Review of Philosophy and Psychology*, 9(2), 213-225.
<https://doi.org/10.1007/s13164-017-0376-x>

- Bernstein, A. (2006). Iva Toguri D'Aquino, 90; « Tokyo Rose » in WWII. *Washington Post*. <https://www.washingtonpost.com/archive/local/2006/09/28/iva-toguri-daquino-90/49fae285-9b55-46c4-96b0-a49f025b9393/>
- Berthelot, B., & Eliakim, P. (2021). Fake news : Comment les théories du complot fragilisent notre économie. *Capital.fr*. <https://www.capital.fr/economie-politique/fake-news-comment-les-theories-du-complot-fragilisent-notre-economie-1400650>
- Beurnez, V. (2022). Piratée, une chaîne d'information ukrainienne diffuse un « deepfake » de Volodymyr Zelensky. *BFMTV*. https://www.bfmtv.com/tech/piratee-une-chaine-d-information-ukrainienne-diffuse-un-deepfake-de-volodymyr-zelensky_AN-202203170296.html
- Biermann, J., de Chantal, L., Korsnes, R., Rohmer, J., & Ündeger, C. (2004). *From Unstructured to Structured Information in Military Intelligence – Some Steps to Improve Information Fusion* (p. 39). NATO. <https://apps.dtic.mil/sti/pdfs/ADA460220.pdf>
- Bièvre-Perrin, F. (2018). *Lampe—Cléopâtre sur un crocodile*. [https://wiki2.org/en/File:67.2.675_Lampe_-_Cleopatre_sur_un_crocodile_\(Mus%C3%A9_de_Die\).jpg](https://wiki2.org/en/File:67.2.675_Lampe_-_Cleopatre_sur_un_crocodile_(Mus%C3%A9_de_Die).jpg)
- Black, S. (2018). The continued influence of retracted witness testimony at jury trials. *Asia Pacific Journal of Advanced Business and Social Studies*, 4(1), 283-290. <https://doi.org/10.25275/apjabssv4i1ss3>
- Bloch, M. (1921). Réflexions d'un historien sur les fausses nouvelles de la guerre. *Revue de synthèse historique*, 33(2), 13-35.
- Bode, L., & Vraga, E. (2021). Correction Experiences on Social Media During COVID-19. *Social Media + Society*, 7(2), 20563051211008829. <https://doi.org/10.1177/20563051211008829>
- Bodenhausen, G. V., Kramer, G. P., & Süsler, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology*, 66(4), 621-632. <https://doi.org/10.1037/0022-3514.66.4.621>
- Bohannon, J. (2014). Replication effort provokes praise—And 'bullying' charges. *Science*, 344(6186), 788-789. <https://doi.org/10.1126/science.344.6186.788>
- Bohner, G., Moskowitz, G., & Chaiken, S. (1995). The interplay of heuristic and systematic processing of social information. *European Review of Social Psychology*, 6(1), 33-68. <https://doi.org/10.1080/14792779443000003>
- Bolsen, T., & Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? *Group Processes & Intergroup Relations*, 21(3), 389-402. <https://doi.org/10.1177/1368430217737855>
- Bor, A., Osmundsen, M., Rasmussen, S. H. R., Bechmann, A., & Petersen, M. B. (2023). « Fact-checking » videos reduce belief in misinformation and improve the quality of news shared on Twitter (a7huq). OSF.io. <https://doi.org/10.31234/osf.io/a7huq>
- Bottemanne, H. (2022). Théories du complot et COVID-19 : Comment naissent les croyances complotistes ? *L'Encéphale*, 48(5), 571-582. <https://doi.org/10.1016/j.encep.2021.12.005>
- Bourguilleau, A. (2023). Jouer pour gagner la guerre. Usages et mésusages du jeu de guerre par les militaires. *DSI (Défense et Sécurité Internationale)*, 164, 82-85.

- Bowles, J., Larreguy, H., & Liu, S. (2020). Countering misinformation via WhatsApp: Preliminary evidence from the COVID-19 pandemic in Zimbabwe. *PLOS ONE*, *15*(10), e0240005. <https://doi.org/10.1371/journal.pone.0240005>
- Boy, D. (2014). Les représentations sociales de la science. In *La science en question(s)* (p. 165-183). Éditions Sciences Humaines. <https://doi.org/10.3917/sh.wievi.2014.01.0165>
- Boy, D. (2021). *Faire confiance à la science ?* (p. 1-6). Sciences po- CEVIPOF. https://www.sciencespo.fr/cevipof/sites/sciencespo.fr/cevipof/files/NoteBaroV12_DB_confiancescience_fevrier2021_VF.pdf
- Boyadjian, J. (2020). Désinformation, non-information ou sur-information ? Les logiques d'exposition à l'actualité en milieux étudiants. *Réseaux*, *222*(4), 21-52. <https://doi.org/10.3917/res.222.0021>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., Grange, J. A., Perugini, M., Spies, J. R., & van 't Veer, A. (2014). The Replication Recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217-224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Brashier, N. M., & Marsh, E. J. (2020). Judging Truth. *Annual Review of Psychology*, *71*(1), 499-515. <https://doi.org/10.1146/annurev-psych-010419-050807>
- Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, *118*(5), 1-3. <https://doi.org/10.1073/pnas.2020043118>
- Brennen, J. S., Simon, F. M., Howard, P. N., & Nielsen, R. K. (2020). *Types, Sources, and Claims of COVID-19 Misinformation* (p. 1-13). University of Oxford.
- Brewster, J., Arvanitis, L., Pavilonis, V., & Wang, M. (2022). Misinformation Monitor : Septembre 2022. *NewsGuard*. <https://www.newsguardtech.com/fr/misinformation-monitor/septembre-2022>
- Bronner, G. (2011). Ce qu'Internet fait à la diffusion des croyances. *Revue européenne des sciences sociales. European Journal of Social Sciences*, *49*(1), Article 49-1. <https://doi.org/10.4000/ress.805>
- Bronner, G. (2013). *La démocratie des crédules*. Presses Universitaires de France. <https://doi.org/10.3917/puf.bron.2013.01>
- Bronner, G. (2021). *Apocalypse cognitive*. Presses Universitaires de France. <https://www.puf.com/apocalypse-cognitive>
- Bronner, G., Cayrol, R., Cordonier, L., Douzet, F., Farinella, R.-M., Favre, A., Garrigues, J., Harfoush, R., Khan, R., Muxel, A., Reichstadt, R., Roder, I., Warusfel, B., & Wieviorka, A. (2022). *Les Lumières à l'ère numérique*. <https://www.elysee.fr/admin/upload/default/0001/12/0f50f46f0941569e780ffc456e62faac59a9e3b7.pdf>
- Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. *Journal of Applied Research in Memory and Cognition*, *8*(1), 108-117. <https://doi.org/10.1016/j.jarmac.2018.09.005>
- Brydges, C. R., Gignac, G. E., & Ecker, U. K. H. (2018). Working memory capacity, short-term memory capacity, and the continued influence effect: A latent-variable analysis. *Intelligence*, *69*, 117-122. <https://doi.org/10.1016/j.intell.2018.03.009>

- Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351-360. [https://doi.org/10.1002/\(sici\)1097-4571\(199106\)42:5<351::AID-ASI5>3.0.CO;2-3](https://doi.org/10.1002/(sici)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3)
- Burnkrant, R. E., & Unnava, H. R. (1989). Self-Referencing: A Strategy for Increasing Processing of Message Content. *Personality and Social Psychology Bulletin*, 15(4), 628-638. <https://doi.org/10.1177/0146167289154015>
- Bushman, B., & Stack, A. (1996). Forbidden Fruit Versus Tainted Fruit: Effects of Warning Labels on Attraction to Television Violence. *Journal of Experimental Psychology: Applied*, 2(3), 207-226. <https://doi.org/10.1037/1076-898X.2.3.207>
- Butler, L. H., Fay, N., & Ecker, U. K. H. (2023). Social endorsement influences the continued belief in corrected misinformation. *Journal of Applied Research in Memory and Cognition*, 12(3), 364-375. <https://doi.org/10.1037/mac0000080>
- Buttrick, N. R., Aczel, B., Aeschbach, L. F., Bakos, B. E., Brühlmann, F., Claypool, H. M., Hüffmeier, J., Kovacs, M., Schuepfer, K., Szecsi, P., Szuts, A., Szöke, O., Thomae, M., Torka, A.-K., Walker, R. J., & Wood, M. J. (2020). Many Labs 5: Registered Replication of Vohs and Schooler (2008), Experiment 1. *Advances in Methods and Practices in Psychological Science*, 3(3), 429-438. <https://doi.org/10.1177/2515245920917931>
- Buzenberg, B. (2014, juin 20). *Finding the truth in 935 lies about war with Iraq*. Center for Public Integrity. <http://publicintegrity.org/inside-publici/finding-the-truth-in-935-lies-about-war-with-iraq/>
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Carnahan, D., Bergan, D. E., & Lee, S. (2021). Do Corrective Effects Last? Results from a Longitudinal Experiment on Beliefs Toward Immigration in the U.S. *Political Behavior*, 43(3), 1227-1246. <https://doi.org/10.1007/s11109-020-09591-9>
- Carpenter, J., Preotiuc-Pietro, D., Clark, J., Flekova, L., Smith, L., Kern, M. L., Buffone, A., Ungar, L., & Seligman, M. (2018). The impact of actively open-minded thinking on social media communication. *Judgment and Decision Making*, 13(6), 562-574. <https://doi.org/10.1017/S1930297500006598>
- Carpenter, S. (2012). Psychology's bold initiative. *Science*, 336(6076), 1558-1561. <https://doi.org/10.1126/science.335.6076.1558>
- Catellani, A. (2021). Changement climatique : Déni, négation et climato-scepticisme. In *Publictionnaire : Dictionnaire encyclopédique et critique des publics*. <https://shorturl.at/3p7XG>
- Cavazos, R. (2019). *The economic cost of bad actors on the Internet* (p. 1-11). CHEQ. https://info.cheq.ai/hubfs/Research/the_economic_cost_fake_news_final.pdf
- CEVIPOF. (2024). *En qu(o)i les Français ont-ils confiance aujourd'hui ? Premiers résultats. Le baromètre de la confiance politique*. CEVIPOF. https://www.sciencespo.fr/cevipof/sites/sciencespo.fr.cevipof/files/BConf_V15_Extraction1_modif.pdf
- Chahed. (2021). Convoi militaire français bloqué au Burkina Faso : Le gouvernement appelle les manifestants au calme. *Anadolu Agency*. <https://www.aa.com.tr/fr/afrique/convoi-militaire-francais-bloqué-au-burkina-faso-le-gouvernement-appelle-les-manifestants-au-calme/2429729>

- Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology*, 39(5), 752-766. <https://doi.org/10.1037/0022-3514.39.5.752>
- Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460-473. <https://doi.org/10.1037/0022-3514.66.3.460>
- Chang, E. P., Ecker, U. K. H., & Page, A. C. (2019). Not wallowing in misery – retractions of negative misinformation are effective in depressive rumination. *Cognition and Emotion*, 33(5), 991-1005. <https://doi.org/10.1080/02699931.2018.1533808>
- Chang, E. Y., Glissmeyer, M., Tonnes, S., Hudson, T., & Johnson, N. (2006). Outcomes of breast cancer in patients who use alternative therapies as primary treatment. *The American Journal of Surgery*, 192(4), 471-473. <https://doi.org/10.1016/j.amjsurg.2006.05.013>
- Chaudet, B. (2009, mars 30). Donnée, information, connaissance. *Logiques processuelles*. <https://brunochoudet.wordpress.com/2009/03/30/donnee-information-connaissance/>
- Chesney, R., & Citron, D. (2019). Deepfakes and the New Disinformation War. *Foreign Affairs*, 98(1), 6.
- Chtourou, A. (2022). Le raisonnement motivé des individus exposés aux informations publicitaires : L'apport de la tâche de sélection de Wason. *Gestion 2000*, 39(2), 81-97. <https://doi.org/10.3917/g2000.392.0081>
- Citrin, D. L., Bloom, D. L., Grutsch, J. F., Mortensen, S. J., & Lis, C. G. (2012). Beliefs and Perceptions of Women with Newly Diagnosed Breast Cancer Who Refused Conventional Treatment in Favor of Alternative Therapies. *The Oncologist*, 17(5), 607-612. <https://doi.org/10.1634/theoncologist.2011-0468>
- Cobb, M. D., Nyhan, B., & Reifler, J. (2013). Beliefs Don't Always Persevere: How Political Figures Are Punished When Positive Information about Them Is Discredited: Beliefs Don't Always Persevere. *Political Psychology*, 34(3), 307-326. <https://doi.org/10.1111/j.1467-9221.2012.00935.x>
- Coder, B. J. (2000). *Q-ships of the Great War*. Air University Press Maxwell AFB.
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self. *Personality and Social Psychology Bulletin*, 26(9), 1151-1164. <https://doi.org/10.1177/01461672002611011>
- Congard, A., Antoine, P., & Gilles, P.-Y. (2012). Assessing the Structural and Psychometric Properties of a New Personality Measure for Use With Military Personnel in the French Armed Forces. *Military Psychology*, 24(3), 289-311. <https://doi.org/10.1080/08995605.2012.678242>
- Connor Desai, S., Pilditch, T. D., & Madsen, J. K. (2020). The rational continued influence of misinformation. *Cognition*, 205(104453). <https://doi.org/10.1016/j.cognition.2020.104453>
- Connor Desai, S., & Reimers, S. (2019). Comparing the use of open and closed questions for Web-based measures of the continued-influence effect. *Behavior Research Methods*, 51(3), 1426-1440. <https://doi.org/10.3758/s13428-018-1066-z>

- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786. <https://doi.org/10.3758/BF03196772>
- Cook, J., & Lewandowsky, S. (2012). *The debunking handbook*. http://www.skepticalscience.com/docs/Debunking_Handbook.pdf
- Cook, J., & Lewandowsky, S. (2016). Rational Irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks. *Topics in Cognitive Science*, *8*(1), 160-179. <https://doi.org/10.1111/tops.12186>
- Cordonier, L., Cafiero, F., & Bronner, G. (2021). Why are conspiracy theories more successful in some countries than in others? An exploratory study on Internet users from 22 Western and non-Western countries. *Social Science Information*, *60*(3), 436-456. <https://doi.org/10.1177/05390184211018961>
- Courchesne, L., Ilhardt, J., & Shapiro, J. N. (2021). Review of social science research on the impact of countermeasures against influence operations. *Harvard Kennedy School Misinformation Review*, *2*(5), 1-17. <https://doi.org/10.37016/mr-2020-79>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- CSA. (2022). *La confiance des Français dans les médias digitaux*. CSA. <https://csa.eu/news/la-confiance-des-francais-dans-les-medias-digitaux/>
- Daneman, M., & Blennerhassett, A. (1984). How to assess the listening comprehension skills of prereaders. *Journal of Educational Psychology*, *76*(6), 1372-1381. <https://doi.org/10.1037/0022-0663.76.6.1372>
- Davidson, D. (1996). The Folly of Trying to Define Truth. *The Journal of Philosophy*, *93*(6), 263-278. <https://doi.org/10.2307/2941075>
- De keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, *65*, 107-110. <https://doi.org/10.1016/j.intell.2017.10.005>
- De Witte Parra, M. (2019). Fortitude, GARBO et le rempart de mensonges. *Leaves*, *8*(1), 1-15. http://dx.doi.org/10.21412/leaves_0806
- Deary, I., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and Educational Achievement. *Intelligence*, *35*, 13-21. <https://doi.org/10.1016/j.intell.2006.02.001>
- Dechêne, A., Stahl, C., Hansen, J., Wänke, M., Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A metaanalytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238-257.
- Delevey, A. (2017, novembre 2). « Fake news » élu mot de l'année par le dictionnaire Collins. *Le Figaro*. <https://shorturl.at/ir62a>
- de Neys, W. (2006). Dual Processing in Reasoning: Two Systems but One Reasoner. *Psychological Science*, *17*(5), 428-433. <https://doi.org/10.1111/j.1467-9280.2006.01723.x>
- de Neys, W., & Verschueren, N. (2006). Working Memory Capacity and a Notorious Brain Teaser: The Case of the Monty Hall Dilemma. *Experimental Psychology*, *53*(2), 123-131. <https://doi.org/10.1027/1618-3169.53.1.123>

- de Saint Laurent, C., Murphy, G., Hegarty, K., & Greene, C. (2022). Measuring the effects of misinformation exposure and beliefs on behavioural intentions: A COVID-19 vaccination study. In *Cognitive Research: Principles and Implications* (Vol. 7, Num. 1, p. 87-106). <https://doi.org/10.1186/s41235-022-00437-y>
- DeSteno, D., Petty, R. E., Wegener, D. T., & Rucker, D. D. (2000). Beyond valence in the perception of likelihood: The role of emotion specificity. *Journal of Personality and Social Psychology*, 78(3), 397-416. <https://doi.org/10.1037/0022-3514.78.3.397>
- Dewar, J. (1998). The Information Age and the Printing Press: Looking Backward to See Ahead. *RAND Corporation*, 1-30. <https://doi.org/10.7249/P8014>
- Dewey, C. (2016). Facebook fake-news writer: 'I think Donald Trump is in the White House because of me'. *Washington Post*. <https://www.washingtonpost.com/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me/>
- Deyries, P. (2022). Les armes de destruction massive de Saddam Hussein. Comprendre l'échec des agences de renseignement. *Institut d'études de géopolitique appliquée*. <https://shorturl.at/k40xY>
- Dieguez, S., & Delouée, S. (2021). *Le complotisme : Cognition, culture, société*. Mardaga.
- Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes? *The International Journal of Press/Politics*, 26(1), 69-91. <https://doi.org/10.1177/1940161220944364>
- Dougherty, M., Thomas, R. P., & Lange, N. D. (2010). Toward an Integrative Theory of Hypothesis Generation, Probability Judgment, and Hypothesis Testing. *Psychology of Learning and Motivation*, 52, 299-342. [https://doi.org/10.1016/S0079-7421\(10\)52008-5](https://doi.org/10.1016/S0079-7421(10)52008-5)
- Drummond, C., Siegrist, M., & Árvai, J. (2020). Limited effects of exposure to fake news about climate change. *Environmental Research Communications*, 2(8), 1-11. <https://doi.org/10.1088/2515-7620/abae77>
- Dunlap, R., & McCright, A. (2008). A Widening Gap: Republican and Democratic Views on Climate Change. *Environment: Science and Policy for Sustainable Development*, 50(5), 26-35. <https://doi.org/10.3200/ENVT.50.5.26-35>
- Dzindolet, M. T., Pierce, L. G., Beck, H. P., Dawe, L. A., & Anderson, B. W. (2001). Predicting Misuse and Disuse of Combat Identification Systems. *Military Psychology*, 13(3), 147-164. https://doi.org/10.1207/S15327876MP1303_2
- Ecker, U. K. H., & Ang, L. C. (2019). Political Attitudes and the Processing of Misinformation Corrections. *Political Psychology*, 40(2), 241-260. <https://doi.org/10.1111/pops.12494>
- Ecker, U. K. H., & Antonio, L. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, 49(4), 631-644. <https://doi.org/10.3758/s13421-020-01129-y>
- Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction? *Journal of Applied Research in Memory and Cognition*, 6(2), 185-192. <https://doi.org/10.1016/j.jarmac.2017.01.014>
- Ecker, U. K. H., Lewandowsky, S., & Apai, J. (2011). Terrorists brought down the plane! No, actually it was a technical fault: Processing corrections of emotive information.

- Quarterly Journal of Experimental Psychology*, 64(2), 283-310.
<https://doi.org/10.1080/17470218.2010.497927>
- Ecker, U. K. H., Lewandowsky, S., & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, 5(41), 1-25.
<https://doi.org/10.1186/s41235-020-00241-6>
- Ecker, U. K. H., Lewandowsky, S., Cheung, C. S. C., & Maybery, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, 85, 101-115.
<https://doi.org/10.1016/j.jml.2015.09.002>
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1, Article 1.
<https://doi.org/10.1038/s44159-021-00006-y>
- Ecker, U. K. H., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, 42(2), 292-304.
<https://doi.org/10.3758/s13421-013-0358-x>
- Ecker, U. K. H., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review*, 18(3), 570-578.
<https://doi.org/10.3758/s13423-011-0065-1>
- Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, 38(8), 1087-1100. <https://doi.org/10.3758/MC.38.8.1087>
- Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language*, 74, 1-15.
<https://doi.org/10.1016/j.jml.2014.03.006>
- Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, 111(1), 36-54.
<https://doi.org/10.1111/bjop.12383>
- Ecker, U. K. H., & Rodricks, A. E. (2020). Do False Allegations Persist? Retracted Misinformation Does Not Continue to Influence Explicit Person Impressions. *Journal of Applied Research in Memory and Cognition*, 9(4), 587-601.
<https://doi.org/10.1016/j.jarmac.2020.08.003>
- Ecker, U. K. H., Sharkey, C. X. M., & Swire-Thompson, B. (2023). Correcting vaccine misinformation: A failure to replicate familiarity or fear-driven backfire effects. *PLOS ONE*, 18(4), e0281140. <https://doi.org/10.1371/journal.pone.0281140>
- Ehrlich, M.-F., & Delafoy, M. (1990). La mémoire de travail : Structure, fonctionnement, capacité. *L'Année psychologique*, 90(3), 403-427.
<https://doi.org/10.3406/psy.1990.29415>
- Einstein, K. L., & Glick, D. M. (2015). Do I Think BLS Data are BS? The Consequences of Conspiracy Theories. *Political Behavior*, 37(3), 679-701.
<https://doi.org/10.1007/s11109-014-9287-z>

- Elizabeth, J. (2014, mai 20). Who are you calling a fact checker? *American Press Institute*.
<https://www.americanpressinstitute.org/fact-checking-project/fact-checker-definition/>
- Ermine, J.-L., Moradi, M., & Brunel, S. (2012). Une chaîne de valeur de la connaissance. *Management international*, 16, 29-40. <https://doi.org/10.7202/1012391ar>
- Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223-241.
<https://doi.org/10.1177/1745691612460685>
- Evon, D. (2022). Bad Deepfake of Zelensky Shared on Ukraine News Site in Reported Hack. *Snopes.Com*. <https://www.snopes.com/news/2022/03/16/zelenskyy-deepfake-shared/>
- Fact-checking at AFP*. (2015, juillet 13). Fact Check. <https://factcheck.afp.com/fact-checking-afp>
- Fang, I. (1997). A summing up. In *A History of Mass Communication: Six Information Revolutions* (1^{re} éd.). Routledge. <https://doi.org/10.4324/9780080508160>
- Farchy, J., & Tallec, S. (2023). De l'information aux industries culturelles, l'hypothèse chahutée de la bulle de filtre. *Questions de communication*, 43, Article 43.
<https://doi.org/10.4000/questionsdecommunication.31474>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Featherstone, J., & Zhang, J. (2020). Feeling angry: The effects of vaccine misinformation and refutational messages on negative emotions and vaccination attitude. *Journal of Health Communication*, 25(9), 692-702.
<https://doi.org/10.1080/10810730.2020.1838671>
- Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the Jury Disregard that Information? The Use of Suspicion to Reduce the Prejudicial Effects of Pretrial Publicity and Inadmissible Testimony. *Personality and Social Psychology Bulletin*, 23(11), 1215-1226. <https://doi.org/10.1177/01461672972311008>
- Fenn, E., Newman, E. J., Pezdek, K., & Garry, M. (2013). The effect of nonprobative photographs on truthiness persists over time. *Acta Psychologica*, 144(1), 207-211.
<https://doi.org/10.1016/j.actpsy.2013.06.004>
- Fletcher, J. (2018). Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance. *Theatre Journal*, 70(4), 455-471.
<https://doi.org/10.1353/tj.2018.0097>
- Floridi, L. (2010). *Information: A Very Short Introduction* (Oxford University Press, USA).
<https://fr.fr1lib.org/book/927105/c16a1c>
- Forgas, J. P., & East, R. (2008). On being happy and gullible: Mood effects on skepticism and the detection of deception. *Journal of Experimental Social Psychology*, 44(5), 1362-1367. <https://doi.org/10.1016/j.jesp.2008.04.010>
- Fox, J. (2019). Confronting Lord Haw-Haw: Rumor and Britain's wartime Anti-Lies Bureau. *The Journal of Modern History*, 91(1), 74-108. <http://dx.doi.org/10.1086/701579>
- François, S. (2016). Désinformation. In *Publictionnaire : Dictionnaire encyclopédique et critique des publics*. <http://publictionnaire.huma-num.fr/wp-content/uploads/2015/09/desinformation.pdf>

- Frápolti, M. J. (2009). Relativism of Truth vs. Dogmatism about Truths a False Dichotomy. *Teorema : Revista Internacional de Filosofía*, 28(3), 65-79. <https://www.jstor.org/stable/43046786>
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25-42. <https://doi.org/10.1257/089533005775196732>
- Freeman, D., Waite, F., Rosebrock, L., Petit, A., Causier, C., East, A., Jenner, L., Teale, A.-L., Carr, L., Mulhall, S., Bold, E., & Lambe, S. (2020). Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychological Medicine*, 52(2), 251-263. <https://doi.org/10.1017/S0033291720001890>
- Furley, P., & Memmert, D. (2012). Working Memory Capacity as Controlled Attention in Tactical Decision Making. *Journal of Sport and Exercise Psychology*, 34(3), 322-344. <https://doi.org/10.1123/jsep.34.3.322>
- Garrett, R. K. (2011). Troubling Consequences of Online Political Rumoring. *Human Communication Research*, 37(2), 255-274. <https://doi.org/10.1111/j.1468-2958.2010.01401.x>
- George, C. (1991). Facilitation in the Wason selection task with a consequent referring to an unsatisfactory outcome. *British Journal of Psychology*, 82(4), 463-472. <https://doi.org/10.1111/j.2044-8295.1991.tb02413.x>
- Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, 52(3), 464-474. <https://doi.org/10.1037/0022-3514.52.3.464>
- Goertzel, T. (2010). Conspiracy theories in science: Conspiracy theories that target specific research can have serious consequences for public health and environmental policies. *EMBO reports*, 11(7), 493-499. <https://doi.org/10.1038/embor.2010.84>
- Goldberg, M., van der Linden, S., Ballew, M., Rosenthal, S., Gustafson, A., & Leiserowitz, A. (2019). The Experience of Consensus: Video as an Effective Medium to Communicate Scientific Agreement on Climate Change. *Science Communication*, 41(5), 659-673. <https://doi.org/10.1177/1075547019874361>
- Gonthier, C., Thomassin, N., & Roulin, J.-L. (2016). The composite complex span: French validation of a short working memory task. *Behavior Research Methods*, 48(1), 233-242. <https://doi.org/10.3758/s13428-015-0566-3>
- Gordon, A., Brooks, J. C. W., Quadflieg, S., Ecker, U. K. H., & Lewandowsky, S. (2017). Exploring the neural substrates of misinformation processing. *Neuropsychologia*, 106, 216-224. <https://doi.org/10.1016/j.neuropsychologia.2017.10.003>
- Graber, D. A. (1990). Seeing is remembering: How visuals contribute to learning from television news. *Journal of Communication*, 40(3), 134-155. <https://doi.org/10.1111/j.1460-2466.1990.tb02275.x>
- Grady, C., & Craik, F. I. M. (2000). Changes in memory processing with age. *Current Opinion in Neurobiology*, 10(2), 224-231. [https://doi.org/10.1016/S0959-4388\(00\)00073-8](https://doi.org/10.1016/S0959-4388(00)00073-8)
- Grant, P. M., Huh, G. A., Perivoliotis, D., Stolar, N. M., & Beck, A. T. (2012). Randomized Trial to Evaluate the Efficacy of Cognitive Therapy for Low-Functioning Patients With Schizophrenia. *Archives of General Psychiatry*, 69(2), 121-127. <https://doi.org/10.1001/archgenpsychiatry.2011.129>

- Graves, L., & Cherubini, F. (2016). *The Rise of Fact-Checking Sites in Europe*. Reuters Institute. <https://reutersinstitute.politics.ox.ac.uk/our-research/rise-fact-checking-sites-europe>
- Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & BREUS, M. (1994). Role of Consciousness and Accessibility of Death-Related Thoughts in Mortality Salience Effects. *Journal of Personality and Social Psychology*, 67, 627-637. <https://doi.org/10.1037//0022-3514.67.4.627>
- Greene, C., Murphy, G., de Saint Laurent, C., Prike, T., Hegarty, K., & Ecker, U. K. H. (2023). Best Practices for Ethical Conduct of Misinformation Research: A Scoping Review and Critical Commentary. *European Psychologist*, 28(3), 139-150. <https://doi.org/10.1027/1016-9040/a000491>
- Greitemeyer, T. (2014). Article retracted, but the message lives on. *Psychonomic Bulletin & Review*, 21(2), 557-561. <https://doi.org/10.3758/s13423-013-0500-6>
- Groh, M., Epstein, Z., Firestone, C., & Picard, R. (2022). Deepfake detection by human crowds, machines, and machine-informed crowds. *Proceedings of the National Academy of Sciences*, 119(1), e2110013119. <https://doi.org/10.1073/pnas.2110013119>
- Guess, A., & Coppock, A. (2020). Does Counter-Attitudinal Information Cause Backlash? Results from Three Large Survey Experiments. *British Journal of Political Science*, 50(4), 1497-1515. <https://doi.org/10.1017/S0007123418000327>
- Guess, A., Lyons, B., Nyhan, B., & Reifler, J. (2018). *Avoiding the echo chamber about echo chambers: Why selective exposure to like-minded political news is less prevalent than you think* (Knight Foundation).
- Guess, A., & Nyhan, B. (2016). *Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign* (p. 1-49). Analysis and Policy Observatory. <https://apo.org.au/node/126961>
- Guillory, J. J., & Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: Remembering facts and believing inferences. *Psychonomic Bulletin & Review*, 17(1), 73-81. <https://doi.org/10.3758/PBR.17.1.73>
- Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, 2(4), 201-209. <https://doi.org/10.1016/j.jarmac.2013.10.001>
- Guillory, J. J., & Geraci, L. (2016). The Persistence of Erroneous Information in Memory: The Effect of Valence on the Acceptance of Corrected Information: Applied Cognitive Psychology. *Applied Cognitive Psychology*, 30(2), 282-288. <https://doi.org/10.1002/acp.3183>
- Gunthert, A. (2008). “Sans retouche” : Histoire d’un mythe photographique. *Études photographiques*, 22, Article 22.
- Gyselinck, V. (1996). Illustrations et modèles mentaux dans la compréhension de textes. *L'Année psychologique*, 96(3), 495-516. <https://doi.org/10.3406/psy.1996.28911>
- Ha, L., Graham, T., & Gray, J. (2022). Where conspiracy theories flourish: A study of YouTube comments and Bill Gates conspiracy theories. *Harvard Kennedy School Misinformation Review*, 3(5), 1-12. <https://doi.org/10.37016/mr-2020-107>

- Hamby, A., Ecker, U. K. H., & Brinberg, D. (2020). How Stories in Memory Perpetuate the Continued Influence of False Information. *Journal of Consumer Psychology, 30*(2), 240-259. <https://doi.org/10.1002/jcpy.1135>
- Hameleers, M., Powell, T., Van Der Meer, T., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication, 37*(2), 281-301. <https://doi.org/10.1080/10584609.2019.1674979>
- Hart, P. S., & Nisbet, E. C. (2012). Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization About Climate Mitigation Policies. *Communication Research, 39*(6), 701-723. <https://doi.org/10.1177/0093650211416646>
- Hartley, S., Dagneaux, S., Palethorpe, C., Londe, V., Liane, M.-T., Decalf, L., Aussert, F., Colas des Francs, C., Leymarie, R., & Royant-Parola, S. (2023). Un programme court de TCCi en groupe par visioconférence dans l'insomnie modifie-t-il les croyances erronées ? *L'Encéphale, 49*(2), 124-129. <https://doi.org/10.1016/j.encep.2022.08.022>
- Houston, A. B. (2018). *Fake news, conspiracy theories, and lies: An information laundering model for homeland security*. Naval Postgraduate School.
- Huchon, T., & Schmidt, J.-B. (2022). *Anti fake news*. First.
- Hughes, H. C., & Waismel-Manor, I. (2021). The Macedonian Fake News Industry and the 2016 US Election. *PS: Political Science & Politics, 54*(1), 19-23. <https://doi.org/10.1017/S1049096520000992>
- Huyghe, F.-B. (2019). *Principales notions sur la stratégie de l'information : Dictionnaire critique*. <https://huyghe.fr/wp-content/uploads/2021/11/47289ed3f2c1e.pdf>
- Ianni, P. (2021). *Burkina : Un convoi militaire français toujours « en attente » près de Ouagadougou* [France 24]. <https://www.france24.com/fr/vid%C3%A9o/20211123-burkina-un-convoi-militaire-fran%C3%A7ais-toujours-en-attente-pr%C3%A8s-de-ouagadougou>
- IFOP. (2023). *Enquête IFOP sur le complotisme et les contre-vérités scientifiques*. IFOP. <https://www.amb-usa.fr/enquete-ifop-complotisme/>
- INSEE. (2020). *Niveau d'éducation de la population – France, portrait social*. INSEE Références. <https://www.insee.fr/fr/statistiques/4797586?sommaire=4928952>
- IPSOS. (2024). *L'impact de la désinformation sur les élections européennes*. IPSOS. <https://www.ipsos.com/sites/default/files/ct/news/documents/2024-03/ipsos-sopra-steria-europeennes-desinformation-rapport-complet.pdf>
- Ithisuphalap, J., Rich, P. R., & Zaragoza, M. S. (2020). Does evaluating belief prior to its retraction influence the efficacy of later corrections? *Memory, 28*(5), 617-631. <https://doi.org/10.1080/09658211.2020.1752731>
- Jansen, V. A. A., Stollenwerk, N., Jensen, H. J., Ramsay, M. E., Edmunds, W. J., & Rhodes, C. J. (2003). Measles Outbreaks in a Population with Declining Vaccine Uptake. *Science, 301*(5634), 804-804. <https://doi.org/10.1126/science.1086726>
- Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139*(1), 39-50. <https://doi.org/10.1016/j.neuroscience.2005.07.002>

- Jeong, S.-H., & Hwang, Y. (2012). Does Multitasking Increase or Decrease Persuasion? Effects of Multitasking on Comprehension and Counterarguing. *Journal of Communication*, 62(4), 571-587. <https://doi.org/10.1111/j.1460-2466.2012.01659.x>
- Jia, L., Shan, J., Xu, G., & Jin, H. (2020). Influence of individual differences in working memory on the continued influence effect of misinformation. *Journal of Cognitive Psychology*, 32(5-6), 494-505. <https://doi.org/10.1080/20445911.2020.1800019>
- Jin, H., Jia, L., Yin, X., Wei, S., & Xu, G. (2022). The Influence of Information Relevance on the Continued Influence Effect of Misinformation. *Journal of Psychiatry and Psychiatric Disorders*, 6(3), 203-218. <https://doi.org/10.26502/jppd.2572-519X0163>
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1420-1436. <https://doi.org/10.1037/0278-7393.20.6.1420>
- Johnson-Laird, P. N., Gawronski, I. B., & Strack, F. (2012). Mental models and consistency. In *Cognitive Consistency: A Unifying Concept in Social Psychology* (Guilford Press).
- Jolley, D., & Douglas, K. M. (2014). The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions. *PLOS ONE*, 9(2), e89177. <https://doi.org/10.1371/journal.pone.0089177>
- Jolley, D., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *British Journal of Social Psychology*, 59(3), 628-640. <https://doi.org/10.1111/bjso.12394>
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407-424. <https://doi.org/10.1017/S1930297500005271>
- Kan, I. P., Pizzonia, K. L., Drummey, A. B., & Mikkelsen, E. J. V. (2021). Exploring factors that mitigate the continued influence of misinformation. *Cognitive Research: Principles and Implications*, 6(1), 1-33. <https://doi.org/10.1186/s41235-021-00335-9>
- Karpf, D. (2019). On Digital Disinformation and Democratic Myths. *Social Science Research Council, Mediawell*, 10. <https://doi.org/10.35650/MD.2012.d.2019>.
- Kassin, S., & Sukel, H. (1997). Coerced Confessions and the Jury: An Experimental Test of the "Harmless Error" Rule. *Law and Human Behavior*, 21(1), 27-46. <https://doi.org/10.1023/A:1024814009769>
- Keil, A. (2018). Book review: « Searching for Lord Haw-Haw: The Political Lives » of William Joyce. *Immigrants & Minorities*, 36(3), 287-289. <https://doi.org/10.1080/02619288.2018.1429153>
- Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013). Updating During Reading Comprehension: Why Causality Matters. *Journal of Experimental Psychology: learning, memory and cognition*, 39(3), 854-865.
- Kendeou, P., Walsh, E., Smith, E., & O'Brien, E. (2014). Knowledge Revision Processes in Refutation Texts. *Discourse Processes*, 51(5-6). <https://doi.org/10.1080/0163853X.2014.913961>
- Kim, K. J., & Sundar, S. S. (2015). Mobile Persuasion: Can Screen Size and Presentation Mode Make a Difference to Trust? *Human Communication Research*, 42(1), 45-70. <https://doi.org/10.1111/hcre.12064>

- Klein, E. (2006). La science dit-elle « LE VRAI » ? *Science & Devenir de l'Homme*, 48, 40-47.
- Klein, R. A., Cook, C. L., Ebersole, C. R., Vitiello, C., Nosek, B. A., Hilgard, J., Ahn, P. H., Brady, A. J., Chartier, C. R., Christopherson, C. D., Clay, S., Collisson, B., Crawford, J. T., Cromar, R., Gardiner, G., Gosnell, C. L., Grahe, J., Hall, C., Howard, I., ... Ratliff, K. A. (2022). Many Labs 4: Failure to Replicate Mortality Salience Effect With and Without Original Author Involvement. *Collabra: Psychology*, 8(1), 35271. <https://doi.org/10.1525/collabra.35271>
- Klen, M. (2016). La désinformation opérationnelle. *Revue de Défense Nationale*, 1(786), 114-118. <https://doi.org/10.3917/rdna.786.0114>
- Koch, A. S., & Forgas, J. P. (2012). Feeling good and feeling truth: The interactive effects of mood and processing fluency on truth judgments. *Journal of Experimental Social Psychology*, 48(2), 481-485. <https://doi.org/10.1016/j.jesp.2011.10.006>
- Kormos, C., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology*, 40, 359-371. <https://doi.org/10.1016/j.jenvp.2014.09.003>
- Kornegay, K. D. (2013). Review of “Double cross: The true story of the D-Day spies”. *Intelligence and National Security*, 28(6), 923-924. <https://doi.org/10.1080/02684527.2012.755064>
- Kraus, F., Bouygues, H. L., & Reichstadt, R. (2023). *La désinformation scientifique des jeunes à l'heure des réseaux sociaux*. Fondation Jean Jaurès. https://www.jean-jaurès.org/publication/la-mesinformation-scientifique-des-jeunes-a-lheure-des-reseaux-sociaux/?post_id=41355&export_pdf=1
- Kubovy, M. (1977). Response Availability and the Apparent Spontaneity of Numerical Choices. *Journal of Experimental Psychology: Human Perception and Performance*, 3(2), 359-364.
- Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the Currency of Democratic Citizenship. *Journal of Politics*, 62(3), 790-816. <https://doi.org/10.1111/0022-3816.00033>
- Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53(4), 636-647. <https://doi.org/10.1037/0022-3514.53.4.636>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lambert, A. J., Khan, S. R., Lickel, B. A., & Fricke, K. (1997). Mood and the correction of positive versus negative stereotypes. *Journal of Personality and Social Psychology*, 72(5), 1002-1016. <https://doi.org/10.1037/0022-3514.72.5.1002>
- Lamy, J., & Petitjean, J. (2021). Mondes de l'écrit et société de l'information à l'époque moderne. *Cahiers d'histoire. Revue d'histoire critique*, 1(149), Article 149. <https://doi.org/10.4000/chrhc.16179>
- Lantian, A., Bagneux, V., Delouée, S., & Gauvrit, N. (2021). Maybe a Free Thinker but Not a Critical One: High Conspiracy Belief Is Associated with Low Critical Thinking Ability. *Applied Cognitive Psychology*, 35(3), 674-684. <https://doi.org/10.1002/acp.3790>

- Larivée, S., Sénéchal, C., & St-Onge, Z. (2018). Le biais de confirmation en clinique. *Enfance*, 4(4), 575-592. <https://doi.org/10.3917/enf2.184.0575>
- Larivée, S., Sénéchal, C., St-Onge, Z., & Sauvé, M.-R. (2019). Le biais de confirmation en recherche. *Revue de psychoéducation*, 48(1), 245-263. <https://doi.org/10.7202/1060013ar>
- Larousse. (2023). Définitions : Information, informations. In *Larousse en ligne*. <https://www.larousse.fr/dictionnaires/francais/information/42993>
- Laurent, V., Kosinski, T., & Rusinek, S. (2023). I know It's false, but I keep thinking as if it were true: A replication study of Johnson and Seifert's (1994) continued influence effect. *Acta Psychologica*, 241, 104094. <https://doi.org/10.1016/j.actpsy.2023.104094>
- Laurent, V., Kosinski, T., & Rusinek, S. (2024). Combating false information in military situations: Striking the right balance between continued influence effect and overcorrection. *Applied Cognitive Psychology*, 38(3), e4214. <https://doi.org/10.1002/acp.4214>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096. <https://doi.org/10.1126/science.aao2998>
- Le Robert. (2023). Définition de information. In *Le Robert, dico en ligne*. <https://dictionnaire.lerobert.com/definition/information>
- Lebret, M. (2015). Des soldats ukrainiens ont reçu un SMS qui prouve que la paix est encore loin. *Slate.fr*. <http://www.slate.fr/story/98037/soldats-debaltseve-texto>
- Lee, D. K. (2016). Alternatives to P value: Confidence interval and effect size. *Korean Journal of Anesthesiology*, 69(6), 555-562. <https://doi.org/10.4097/kjae.2016.69.6.555>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, 13(3), 106-131. <https://doi.org/10.1177/1529100612451018>
- Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for Fact, Fiction, and Misinformation: The Iraq War 2003. *Psychological Science*, 16(3), 190-195. <https://doi.org/10.1111/j.0956-7976.2005.00802.x>
- Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2009). Misinformation and the "War on Terror": When memory turns fiction into fact. In W. G. K. Stritzke, S. Lewandowsky, D. Denemark, J. Clare, & F. Morgan (Éds.), *Terrorism and Torture* (p. 179-203). Cambridge University Press. <https://doi.org/10.1017/CBO9780511581199.010>
- Li, H. O.-Y., Bailey, A., Huynh, D., & Chan, J. (2020). YouTube as a source of information on COVID-19: A pandemic of misinformation? *BMJ Global Health*, 5(5), e002604. <https://doi.org/10.1136/bmjgh-2020-002604>
- Liu, Q., & Nesbit, J. C. (2024). The Relation Between Need for Cognition and Academic Achievement: A Meta-Analysis. *Review of Educational Research*, 94(2), 155-192. <https://doi.org/10.3102/00346543231160474>
- Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and

- USA. *Nature Human Behaviour*, 5(3), 337-348. <https://doi.org/10.1038/s41562-021-01056-1>
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, 37(11), 2098-2109. <https://doi.org/10.1037/0022-3514.37.11.2098>
- Ludwig, P. (2016). Vérité (A). In *L'encyclopédie philosophique*. <https://encyclo-philo.fr/verite-a>
- Luengo, M., & García-Marín, D. (2020). The performance of truth: Politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation. *American Journal of Cultural Sociology*, 8(3), 405-427. <https://doi.org/10.1057/s41290-020-00115-w>
- MacFarlane, D., Hurlstone, M. J., & Ecker, U. K. H. (2018). Reducing demand for ineffective health remedies: Overcoming the illusion of causality. *Psychology & Health*, 33(12), 1472-1489. <https://doi.org/10.1080/08870446.2018.1508685>
- Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The Other Half of the Story: Effect Size Analysis in Quantitative Research. *CBE—Life Sciences Education*, 12(3), 345-351. <https://doi.org/10.1187/cbe.13-04-0082>
- Majerus, S., & Poncelet, M. (2017). Dyslexie et déficits de la mémoire à court terme/de travail : Implications pour la remédiation. *ANAE: Approche Neuropsychologique des Apprentissages chez l'Enfant*, 148, 295-302. <https://hdl.handle.net/2268/217880>
- Mandjo, S. (2021). Non, ces photos ne montrent pas des soldats étrangers pillant de l'or au Mali. *AFP Factuel*. <https://factuel.afp.com/non-ces-photos-ne-montrent-pas-des-soldats-etrangers-pillant-de-lor-au-mali>
- Manovich, L. (2011). Inside Photoshop. *Computational Culture*, 1. <http://computationalculture.net/inside-photoshop/>
- Marsh, E. J., & Stanley, M. L. (2021). False beliefs: By products of an adaptive knowledge base? In R. Greifeneder, M. E. Jaffé, E. J. Newman, & N. Schwarz (Éds.), *The Psychology of Fake News: Accepting, Sharing and Correcting Misinformation* (Taylor&Francis, p. 144-159). Routledge. <https://library.oapen.org/bitstream/handle/20.500.12657/46921/9781000179033.pdf?sque#page=144>
- Martin, L., Seta, J., & Crelia, R. (1990). Assimilation and Contrast as a Function of People's Willingness and Ability to Expend Effort in Forming an Impression. *Journal of Personality and Social Psychology*, 59(1), 27-37. <https://doi.org/10.1037/0022-3514.59.1.27>
- Martins, D., & Le Bouëdec, B. (1998). La production d'inférences lors de la compréhension de textes chez des adultes : Une analyse de la littérature. *L'Année psychologique*, 98(3), 511-543. <https://doi.org/10.3406/psy.1998.28581>
- Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, 75(2), 314-318. <https://doi.org/10.1037/0022-0663.75.2.314>
- Mayers, K. (2019). *Anti-vaxxers are costing Americans billions each year*. <https://finance.yahoo.com/news/antivaxxers-costing-americans-billions-each-year-191839191.html>

- McHoskey, J. W. (1995). Case Closed? On the John F. Kennedy Assassination: Biased Assimilation of Evidence and Attitude Polarization. *Basic and Applied Social Psychology*, 17(3), 395-409. https://doi.org/10.1207/s15324834basp1703_7
- McIlhiney, P., Gignac, G. E., Ecker, U. K. H., Kennedy, B. L., & Weinborn, M. (2023). Executive function and the continued influence of misinformation: A latent-variable analysis. *PLOS ONE*, 18(4), e0283951. <https://doi.org/10.1371/journal.pone.0283951>
- McIlhiney, P., Gignac, G. E., Ecker, U. K. H., Kennedy, B., & Weinborn, M. (2022). *Executive Function and the Continued Influence of Misinformation: A Latent Variable Analysis*. PsyArXiv. <https://doi.org/10.31234/osf.io/pgsyd>
- McIlhiney, P., Gignac, G. E., Weinborn, M., & Ecker, U. K. H. (2022). Sensitivity to misinformation retractions in the continued influence paradigm: Evidence for stability. *Quarterly Journal of Experimental Psychology*, 75(7), 1259-1271. <https://doi.org/10.1177/17470218211048986>
- McKey, M. (2014). How Aviation Increased the Effectiveness of Psychological Warfare in the War with Japan During World War II. *Aviation in America*. <https://sites.google.com/site/aviationinamerica/home/how-aviation-increased-the-effectiveness-of-psychological-warfare-in-the-war-with-japan-during-world-war-ii>
- Meckel, M., & Steinacker, L. (2021). Hybrid Reality: The Rise of Deepfakes and Diverging Truths. *Morals & Machines*, 1(1), 10-21. <https://doi.org/10.5771/2747-5174-2021-1-10>
- Melki, J., Tamim, H., Hadid, D., Makki, M., Amine, J. E., & Hitti, E. (2021). Mitigating infodemics: The relationship between news exposure and trust and belief in COVID-19 fake news and social media spreading. *PLOS ONE*, 16(6), e0252830. <https://doi.org/10.1371/journal.pone.0252830>
- Mille, A. (2014). D'Internet au web. In M. E. Sinatra & M. Vitali-Rosati, *Pratiques de l'édition numérique* (Presses de l'Université de Montréal). <http://parcoursnumeriques-pum.ca/1-pratiques/chapitre2.html>
- Miller, A. L., Wissman, K. T., & Peterson, D. J. (2022). The continued influence effect: Examining how age, retraction, and delay impact inferential reasoning. *Applied Cognitive Psychology*, 36(3), 708-723. <https://doi.org/10.1002/acp.3939>
- Misra, S. (1992). Is Conventional Debriefing Adequate? An Ethical Issue in Consumer Research. *Journal of the Academy of Marketing Science*, 20(3), 269-273. <https://doi.org/10.1177/0092070392203009>
- Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*, 65(2), 180-212. <https://doi.org/10.1177/0002764219878224>
- Mosleh, M., Martel, C., Eckles, D., & Rand, D. (2021). Perverse Downstream Consequences of Debunking: Being Corrected by Another User for Posting False Political News Increases Subsequent Sharing of Low Quality, Partisan, and Toxic Content in a Twitter Field Experiment. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-13. <https://doi.org/10.1145/3411764.3445642>
- Mosleh, M., Pennycook, G., & Rand, D. G. (2022). Field Experiments on Social Media. *Current Directions in Psychological Science*, 31(1), 69-75. <https://doi.org/10.1177/09637214211054761>

- Munro, G. D. (2010). The Scientific Impotence Excuse: Discounting Belief-Threatening Scientific Abstracts. *Journal of Applied Social Psychology, 40*(3), 579-600. <https://doi.org/10.1111/j.1559-1816.2010.00588.x>
- Murphy, G., Laurent, C. de S., Reynolds, M., Aftab, O., Hegarty, K., Sun, Y., & Greene, C. M. (2023). What do we study when we study misinformation? A scoping review of experimental research (2016-2022). *Harvard Kennedy School Misinformation Review, 4*(6), 1-57. <https://doi.org/10.37016/mr-2020-130>
- Myers, J., & O'Brien, E. (1998). Accessing the discourse representation while reading. *Discourse Processes, 26*(2-3), 131-157. <https://doi.org/10.1080/01638539809545042>
- NATO. (2020). *NATO's approach to countering disinformation: A focus on COVID-19*. NATO. <https://www.nato.int/cps/en/natohq/177273.htm>
- NATO. (2022). *Setting the record straight: De-bunking Russian disinformation on NATO*. NATO. <https://www.nato.int/cps/en/natohq/115204.htm>
- Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review, 19*(5), 969-974. <https://doi.org/10.3758/s13423-012-0292-0>
- Nosek, B. A. (2014). *Data from Investigating Variation in Replicability: A "Many Labs" Replication Project* (1). 2(1), Article 1. <https://doi.org/10.5334/jopd.ad>
- Novin, A., & Meyers, E. (2017). Making Sense of Conflicting Science Information: Exploring Bias in the Search Engine Result Page. *CHIIR 17: Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, 1*(1), 175-184. <https://doi.org/10.1145/3020165.3020185>
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2020). Taking Fact-Checks Literally But Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability. *Political Behavior, 42*(3), 939-960. <https://doi.org/10.1007/s11109-019-09528-x>
- Nyhan, B., & Reifler, J. (2010). When Corrections Fail: The Persistence of Political misperceptions. *Political Behavior, 32*(2), 303-330. <https://doi.org/10.1007/s11109-010-9112-2>
- Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine, 33*(3), 459-464. <https://doi.org/10.1016/j.vaccine.2014.11.017>
- Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective Messages in Vaccine Promotion: A Randomized Trial. *Pediatrics, 133*(4), 835-842. <https://doi.org/10.1542/peds.2013-2365>
- Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The Hazards of Correcting Myths About Health Care Reform. *Medical Care, 51*(2), 127-132. <https://doi.org/10.1097/MLR.0b013e318279486b>
- O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1200-1210. <https://doi.org/10.1037/0278-7393.24.5.1200>

- Open Science Collaboration. (2012). An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science*, 7(6), 657-660. <https://doi.org/10.1177/1745691612462588>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- O'Rear, A. E., & Radvansky, G. A. (2020). Failure to accept retractions: A contribution to the continued influence effect. *Memory & Cognition*, 48(1), 127-144. <https://doi.org/10.3758/s13421-019-00967-9>
- Parly, F. (2021). *Présentation de la doctrine militaire de lutte informatique d'influence* [Discours]. https://www.defense.gouv.fr/sites/default/files/ema/discours_de_florence_parly_la_doctrine_de_lutte_informatique_dinfluence.pdf
- Pashler, H., & Harris, C. R. (2012). Is the Replicability Crisis Overblown? Three Arguments Examined. *Perspectives on Psychological Science*, 7(6), 531-536. <https://doi.org/10.1177/1745691612463401>
- Pashler, H., & Wagenmakers, E. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, 7(6), 528-530. <https://doi.org/10.1177/1745691612465253>
- Pasquinelli, E., & Bronner, G. (2021). *Eduquer à l'esprit critique*. Conseil Scientifique du Ministère de l'Éducation nationale. https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conseil_scientifique_education_nationale/Resources_pedagogiques/VDEF_Eduquer_a_lesprit_critique_CSEN.pdf
- Patel, N. (2018). *The Cognitive Reflection Test: A Measure of Intuition/Reflection, Numeracy, and Insight Problem Solving, and the Implications for Understanding Real-World Judgments and Beliefs* [University of Missouri-Columbia]. <https://doi.org/10.31234/osf.io/xeyj8>
- Paynter, J., Luskin-Saxby, S., Keen, D., Fordyce, K., Frost, G., Imms, C., Miller, S., Trembath, D., Tucker, M., & Ecker, U. (2019). Evaluation of a template for countering misinformation—Real-world Autism treatment myth debunking. *PLOS ONE*, 14(1), e0210746. <https://doi.org/10.1371/journal.pone.0210746>
- Pehlivanoglu, D., Lin, T., Deceus, F., Heemskerk, A., Ebner, N. C., & Cahill, B. S. (2021). The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cognitive Research: Principles and Implications*, 6(24), 1-13. <https://doi.org/10.1186/s41235-021-00292-3>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology. General*, 147(12), 1865-1880. <https://doi.org/10.1037/xge0000465>
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, 48(1), 341-348. <https://doi.org/10.3758/s13428-015-0576-1>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), Article 7855. <https://doi.org/10.1038/s41586-021-03344-2>

- Pennycook, G., Fugelsang, J., & Koehler, D. (2015). Everyday Consequences of Analytic Thinking. *Current Directions in Psychological Science*, 24(6), 425-432. <https://doi.org/10.1177/09637214155604610>
- Petty, R. E., Barden, J., & Wheeler, S. C. (2009). The Elaboration Likelihood Model of persuasion: Developing health promotions for sustained behavioral change. In *Emerging theories in health promotion practice and research* (R. J. DiClemente, R. A. Crosby, M. Kegler, p. 185-214). Jossey-Bass/Wiley. <https://psycnet.apa.org/record/2009-19878-007>
- Petty, R. E., Brinol, P., Loersch, C., & MacCaslin, M. J. (2009). The Need for Cognition. In *Handbook of individual differences in social behavior* (M. R. Leary & R. H. Hoyle (Eds.), p. 318-329). Guilford Press. <https://psycnet.apa.org/record/2009-12071-021>
- Petty, R. E., & Cacioppo, J. T. (1981). Issue Involvement As a Moderator of the Effects on Attitude of Advertising Content and Context. *Advances in Consumer Research*, 8(1), 20-24.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, 19, 123-205. [https://doi.org/10.1016/S0065-2601\(08\)60214-2](https://doi.org/10.1016/S0065-2601(08)60214-2)
- Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal Involvement as a Determinant of Argument-Based Persuasion. *Journal of Personality and Social Psychology*, 41(5), 845-857. <https://doi.org/10.1037/0022-3514.41.5.847>
- Planetoscope—Statistiques : Publications d'articles scientifiques dans le monde.* (s. d.). Consulté 9 janvier 2023, à l'adresse <https://www.planetoscope.com/entreprises/2026-publications-d-articles-scientifiques-dans-le-monde.html>
- Pluviano, S., Watt, C., & Sala, S. D. (2017). Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLOS ONE*, 12(7), e0181640. <https://doi.org/10.1371/journal.pone.0181640>
- Poirier, L. (2012). *Névrosisme et bien-être chez les nouveaux parents : Une question de perception des conflits travail-famille ?* [Université de Moncton]. <https://library-archives.canada.ca/eng/services/services-libraries/theses/Pages/item.aspx?idNumber=812063874>
- Poland, G. A., & Spier, R. (2010). Fear, misinformation, and innumerates: How the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine*, 28(12), 2361-2362. <https://doi.org/10.1016/j.vaccine.2010.02.052>
- Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences*, 118(37), e2104235118. <https://doi.org/10.1073/pnas.2104235118>
- Porter, E., & Wood, T. J. (2022). Political Misinformation and Factual Corrections on the Facebook News Feed: Experimental Evidence. *The Journal of Politics*, 84(3), 1812-1817. <https://doi.org/10.1086/719271>
- Porter, E., Wood, T. J., & Bahador, B. (2019). Can presidential misinformation on climate change be corrected? Evidence from Internet and phone experiments. *Research & Politics*, 6(3). <https://doi.org/10.1177/2053168019864784>

- Porter, E., Wood, T. J., & Kirby, D. (2018). Sex Trafficking, Russian Infiltration, Birth Certificates, and Pedophilia: A Survey Experiment Correcting Fake News. *Journal of Experimental Political Science*, 5(2), 159-164. <https://doi.org/10.1017/XPS.2017.32>
- Posetti, J., & Matthews, A. (2018). *A short guide to the history of 'fake news' and disinformation*. International Center for Journalists. <https://shorturl.at/738R5>
- Prasad, M., Perrin, A. J., Bezila, K., Hoffman, S. G., Kindleberger, K., Manturuk, K., & Powers, A. S. (2009). "There Must Be a Reason": Osama, Saddam, and Inferred Justification. *Sociological Inquiry*, 79(2), 142-162. <https://doi.org/10.1111/j.1475-682X.2009.00280.x>
- Priester, J., Wegener, D., Petty, R., & Fabrigar, L. (1999). Examining the Psychological Process Underlying the Sleeper Effect: The Elaboration Likelihood Model Explanation. *Media Psychology*, 1(1), 27-48. https://doi.org/10.1207/s1532785xmep0101_3
- Prior, M., Sood, G., & Khanna, K. (2015). You Cannot be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions. *Quarterly Journal of Political Science*, 10(4), 489-518. <https://doi.org/10.1561/100.00014127>
- Ranney, M. A., & Clark, D. (2016). Climate Change Conceptual Change: Scientific Information Can Transform Attitudes. *Topics in Cognitive Science*, 8(1), 49-75. <https://doi.org/10.1111/tops.12187>
- Rapp, D. N. (2016). The Consequences of Reading Inaccurate Information. *Current Directions in Psychological Science*, 25(4), 281-285. <https://doi.org/10.1177/0963721416649347>
- Rapp, D. N., Hinze, S. R., Slaten, D. G., & Horton, W. S. (2014). Amazing Stories: Acquiring and Avoiding Inaccurate Information From Fiction. *Discourse Processes*, 51(1-2), 50-74. <https://doi.org/10.1080/0163853X.2013.855048>
- Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, 35(8), 2019-2032. <https://doi.org/10.3758/BF03192934>
- Rapp, D. N., & van der Broek, P. (2005). Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science*, 14(5), 276-279. <https://doi.org/10.1111/j.0963-7214.2005.00380.x>
- Ratzan, S. C. (2010). Setting the Record Straight: Vaccines, Autism, and the *Lancet*. *Journal of Health Communication*, 15(3), 237-239. <https://doi.org/10.1080/10810731003780714>
- Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring Working Memory Capacity with Automated Complex Span Tasks. *European Journal of Psychological Assessment*, 28(3), 164-171. <https://doi.org/10.1027/1015-5759/a000123>
- Reichstadt, R. (2019). Enquête complotisme 2019 : Les grands enseignements. *Fondation Jean-Jaurès*. <https://www.jean-jaures.org/publication/enquete-complotisme-2019-les-grands-enseignements/>
- Reichstadt, R., & Fourquet, J. (2020). L'épidémie dans l'épidémie : Thèses complotistes et Covid-19. *Fondation Jean-Jaurès*. <https://www.jean-jaures.org/publication/lepidemie-dans-lepidemie-theses-complotistes-et-covid-19/>

- Reichstadt, R., & Igounet, V. (2018). *Complotisme et négationnisme, un panorama*. Conspiracy Watch. <https://www.fondationshoah.org/sites/default/files/1-antisemitisme/conspiracywatch-rapport2018-panorama.pdf>
- Reuters Fact Check. (2021, novembre 16). Fact Check-Clip of Jacinda Ardern smoking cannabis is a deepfake. *Reuters*. <https://www.reuters.com/article/factcheck-newzealand-ardern-idUSL1N2S71BZ>
- Reuters Fact Check. (2022, mars 17). Fact Check-Doctored video appears to show Putin announcing peace. *Reuters*. <https://www.reuters.com/article/factcheck-putin-address-idUSL2N2VK1CC>
- Revault d'Allonnes, M. (2018). *La faiblesse du vrai : Ce que la post-vérité fait à notre monde commun* (Seuil).
- Rich, P. R., Donovan, A. M., & Rapp, D. N. (2022). Cause typicality and the continued influence effect. *Journal of Experimental Psychology: Applied*, 29(2), 221-238. <https://doi.org/10.1037/xap0000454>
- Rich, P. R., & Zaragoza, M. S. (2015). The Continued Influence of Implied and Explicitly Stated Misinformation in News Reports. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 42(1), 62-74. <https://doi.org/10.1037/xlm0000155>
- Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., & Wilson, C. (2018). Auditing Partisan Audience Bias within Google Search. *Proceedings of the ACM on Human-Computer Interaction*, 2(148), 1-22. <https://doi.org/10.1145/3274417>
- Rochlin, G. I. (1991). Iran Air Flight 655 and the USS Vincennes. In T. R. Porte (Éd.), *Social Responses to Large Technical Systems* (Todd R. Porte, p. 99-125). Springer Netherlands. https://doi.org/10.1007/978-94-011-3400-2_5
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14(11), e12633. <https://doi.org/10.1111/phc3.12633>
- Roozenbeek, J., Schneider, C., Dryhurst, S., Kerr, J., Freeman, A., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, 7(10), 201199. <https://doi.org/10.1098/rsos.201199>
- Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34), eabo6254. <https://doi.org/10.1126/sciadv.abo6254>
- Roozenbeek, J., van der Linden, S., & Nygren, T. (2020). Prebunking interventions based on “inoculation” theory can reduce susceptibility to misinformation across cultures. *Harvard Kennedy School Misinformation Review*, 1(2), 1-23. <https://doi.org/10.37016//mr-2020-008>
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm. *Journal of Personality and Social Psychology*, 32(5), 880-892. <https://doi.org/10.1037/0022-3514.32.5.880>
- Ross, R. M., Rand, D. G., & Pennycook, G. (2021). Beyond “fake news”: Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgement and Decision Making*, 16(2), 484-504. <https://doi.org/10.31234/osf.io/cgsx6>

- Rothman, A. J., & Schwarz, N. (1998). Constructing Perceptions of Vulnerability: Personal Relevance and the Use of Experiential Information in Health Judgments. *Personality and Social Psychology Bulletin*, 24(10), 1053-1064. <https://doi.org/10.1177/01461672982410003>
- Rouse, E. (s. d.). History of PSYOP. *Psywarrior*. Consulté 25 mars 2022, à l'adresse <http://www.psywarrior.com/psyhist.html>
- Roussin, J. (2021). Fausses nouvelles : Trouble dans la croyance. *Esprit*, 479(11), 51-63. <https://doi.org/10.3917/espri.2111.0051>
- Ruder, M., & Bless, H. (2003). Mood and the reliance on the ease of retrieval heuristic. *Journal of Personality and Social Psychology*, 85(1), 20-32. <https://doi.org/10.1037/0022-3514.85.1.20>
- Ryan, E. (2022, avril 19). From Sofia to Springfield, fact-checking extends its reach. *Duke Reporters' Lab*. <https://reporterslab.org/from-sofia-to-springfield-fact-checking-extends-its-reach/>
- Ryckmans, G., Carton, A., & Payen, H. (2022). Ce faux discours de Vladimir Poutine annonçant « la paix » avec l'Ukraine est un deepfake. *EDMO Belux*. <https://belux.edmo.eu/fr/ce-faux-discours-de-vladimir-poutine-annonant-la-paix-avec-lukraine-est-un-deepfake/>
- Saklofske, D. H., Kelly, I. W., & Janzen, B. L. (1995). Neuroticism, depression, and depression proneness. *Personality and Individual Differences*, 18(1), 27-31. [https://doi.org/10.1016/0191-8869\(94\)00128-F](https://doi.org/10.1016/0191-8869(94)00128-F)
- Sanders, L. (2020). *The difference between what Republicans and Democrats believe to be true about COVID-19*. Yahoo News/YouGov. <https://today.yougov.com/topics/politics/articles-reports/2020/05/26/republicans-democrats-misinformation>
- Sanderson, J. A., Bowden, V., Swire-Thompson, B., Lewandowsky, S., & Ecker, U. K. H. (2023). Listening to misinformation while driving: Cognitive load and the effectiveness of (repeated) corrections. *Journal of Applied Research in Memory and Cognition*, 12(3), 325-334. <https://doi.org/10.1037/mac0000057>
- Sanderson, J. A., Farrell, S., & Ecker, U. K. H. (2022). Examining the role of information integration in the continued influence effect using an event segmentation approach. *PLOS ONE*, 17(7), e0271566. <https://doi.org/10.1371/journal.pone.0271566>
- Sanderson, J. A., Gignac, G. E., & Ecker, U. K. H. (2021). Working memory capacity, removal efficiency and event specific memory as predictors of misinformation reliance. *Journal of Cognitive Psychology*, 33(5), 518-532. <https://doi.org/10.1080/20445911.2021.1931243>
- Saurez-vous répondre à ce problème posé aux étudiants de Harvard ? (2018, septembre 10). *Gentside*. https://www.gentside.com/insolite/saurez-vous-repondre-a-ce-probleme-pose-aux-etudiants-de-harvard_art47028.html
- Sauvayre, R. (2021). Ethics of belief, trust and epistemic value. The case of the scientific controversy surrounding the measles vaccine. *Journal of Leadership Accountability and Ethics*, 18(4), 24-34. <https://doi.org/10.33423/jlae.v18i4.4607>
- Savolainen, R., & Kari, J. (2004). Placing the Internet in information source horizons. A study of information seeking by Internet users in the context of self-development. *Library &*

- Information Science Research*, 26(4), 415-433.
<https://doi.org/10.1016/j.lisr.2004.04.004>
- Schudson, M. (2004). Notes on Scandal and the Watergate Legacy. *American Behavioral Scientist*, 47(9), 1231-1238. <https://doi.org/10.1177/0002764203262345>
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science*, 18(5), 429-434. <https://doi.org/10.1111/j.1467-9280.2007.01917.x>
- Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? In *Psychology of Learning and Motivation* (Vol. 41, p. 265-292). Elsevier. [https://doi.org/10.1016/S0079-7421\(02\)80009-3](https://doi.org/10.1016/S0079-7421(02)80009-3)
- Sénéchal, C., Larivée, S., Richard, E., & Robert, Y. (2004). Vaccin RRO et autisme : La désinformation à l'œuvre. *Revue de psychoéducation*, 33(1), 205-227.
- Sharma, M., Yadav, K., Yadav, N., & Ferdinand, K. C. (2017). Zika virus pandemic— Analysis of Facebook as a social media health information platform. *American Journal of Infection Control*, 45(3), 301-302.
<https://doi.org/10.1016/j.ajic.2016.08.022>
- Sherman, D. K., & Cohen, G. L. (2002). Accepting Threatening Information: Self-Affirmation and the Reduction of Defensive Biases. *Current Directions in Psychological Science*, 11(4), 119-123. <https://doi.org/10.1111/1467-8721.00182>
- Simis, M. J., Madden, H., Cacciatore, M. A., & Yeo, S. K. (2016). The lure of rationality: Why does the deficit model persist in science communication? *Public Understanding of Science*, 25(4), 400-414. <https://doi.org/10.1177/0963662516629749>
- Simmons, J., Nelson, L., & Simonsohn, U. (2012). A 21 words solution. *The Official Newsletter of the Society for Personality and Social Psychology*, 26(2), 4-7.
<http://dx.doi.org/10.2139/ssrn.2160588>
- Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science*, 9(1), 76-80. <https://doi.org/10.1177/1745691613514755>
- Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How Warnings about False Claims Become Recommendations. *Journal of Consumer Research*, 31(4), 713-724.
<https://doi.org/10.1086/426605>
- Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2), 143-153.
<https://doi.org/10.1016/j.obhdp.2006.01.005>
- Sommers, S. R., & Kassir, S. M. (2001). On the Many Impacts of Inadmissible Testimony: Selective Compliance, Need for Cognition, and the Overcorrection Bias. *Personality and Social Psychology Bulletin*, 27(10), 1368-1377.
<https://doi.org/10.1177/01461672012710012>
- Stagnaro, M. N., Ross, R. M., Pennycook, G., & Rand, D. G. (2019). Cross-cultural support for a link between analytic thinking and disbelief in God: Evidence from India and the United Kingdom. *Judgment and Decision Making*, 14(2), 179-186.
<https://doi.org/10.1017/S1930297500003417>

- Ståhl, T., & van Prooijen, J.-W. (2018). Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. *Personality and Individual Differences, 122*, 155-163. <https://doi.org/10.1016/j.paid.2017.10.026>
- Stencel, M., & Griffin, R. (2018, février 22). Fact-checking triples over four years. *Duke Reporters' Lab*. <https://reporterslab.org/fact-checking-triples-over-four-years/>
- Stencel, M., Ryan, E., & Luther, J. (2022, juin 17). Fact-checkers extend their global reach with 391 outlets, but growth has slowed. *Duke Reporters' Lab*. <https://reporterslab.org/fact-checkers-extend-their-global-reach-with-391-outlets-but-growth-has-slowed/>
- Stencel, M., Ryan, E., & Luther, J. (2023, juin 21). Misinformation spreads, but fact-checking has leveled off. *Duke Reporters' Lab*. <https://reporterslab.org/misinformation-spreads-but-fact-checking-has-leveled-off/>
- Stocking, G., van Kessel, P., Barthel, M., Eva Matsa, K., & Khuzam, M. (2020, septembre 28). Many Americans Get News on YouTube, Where News Organizations and Independent Producers Thrive Side by Side. *Pew Research Center's Journalism Project*. <https://www.pewresearch.org/journalism/2020/09/28/many-americans-get-news-on-youtube-where-news-organizations-and-independent-producers-thrive-side-by-side/>
- Stover, D. (2018). Garlin Gilchrist: Fighting fake news and the information apocalypse. *Bulletin of the Atomic Scientists, 74*(4), 283-288. <https://doi.org/10.1080/00963402.2018.1486618>
- Subramanian, S. (2017). Meet the Macedonian Teens Who Mastered Fake News and Corrupted the US Election. *Wired*. <https://www.wired.com/2017/02/veles-macedonia-fake-news/>
- Sundar, S. S. (2000). Multimedia Effects on Processing and Perception of Online News: A Study of Picture, Audio, and Video Downloads. *Journalism & Mass Communication Quarterly, 77*(3), 480-499. <https://doi.org/10.1177/107769900007700302>
- Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing Is Believing: Is Video Modality More Powerful in Spreading Fake News via Online Messaging Apps? *Journal of Computer-Mediated Communication, 26*(6), 301-319. <https://doi.org/10.1093/jcmc/zmab010>
- Susmann, M. W., & Wegener, D. (2021). The role of discomfort in the continued influence effect of misinformation. *Memory & Cognition, 50*(2), 435-448. <https://doi.org/10.3758/s13421-021-01232-8>
- Svedholm-Häkkinen, A. M., & Lindeman, M. (2018). Actively open-minded thinking: Development of a shortened scale and disentangling attitudes towards knowledge and people. *Thinking & Reasoning, 24*(1), 21-40. <https://doi.org/10.1080/13546783.2017.1378723>
- Swami, V., Barron, D., Weis, L., & Furnham, A. (2018). To Brexit or not to Brexit: The roles of Islamophobia, conspiracist beliefs, and integrated threat in voting intentions for the United Kingdom European Union membership referendum. *British Journal of Psychology, 109*(1), 156-179. <https://doi.org/10.1111/bjop.12252>
- Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition, 133*(3), 572-585. <https://doi.org/10.1016/j.cognition.2014.08.006>

- Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), 1948-1961. <https://doi.org/10.1037/xlm0000422>
- Swire-Thompson, B., Cook, J., Butler, L. H., Sanderson, J. A., Lewandowsky, S., & Ecker, U. K. H. (2021). Correction format has a limited role when debunking misinformation. *Cognitive Research: Principles and Implications*, 6(1), 83-99. <https://doi.org/10.1186/s41235-021-00346-6>
- Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the Backfire Effect: Measurement and Design Considerations. *Journal of Applied Research in Memory and Cognition*, 9(3), 286-299. <https://doi.org/10.1016/j.jarmac.2020.06.006>
- Tal-Or, N., & Cohen, J. (2010). Understanding audience involvement: Conceptualizing and manipulating identification and transportation. *Poetics*, 38(4), 402-418. <https://doi.org/10.1016/j.poetic.2010.05.004>
- Tavares, E. S. (2001). *Operation Fortitude: The closed loop D-Day deception plan*. [Air Command and Staff College; Air University Maxwell]. <https://apps.dtic.mil/sti/pdfs/ADA407763.pdf>
- Ternovski, J., Kalla, J., & Aronow, P. (2022). The Negative Consequences of Informing Voters about Deepfakes: Evidence from Two Survey Experiments. *Journal of Online Trust and Safety*, 1(2), Article 2. <https://doi.org/10.54501/jots.v1i2.28>
- Thompson-Fuller, T. (2021, octobre 13). Attention à ce deepfake sur la Première ministre néo-zélandaise Jacinda Ardern. *AFP Factuel*. <https://factuel.afp.com/http%253A%252F%252Fdoc.afp.com%252F9PL6EU-1>
- Thorson, E. (2016). Belief Echoes: The Persistent Effects of Corrected Misinformation. *Political Communication*, 33(3), 460-480. <https://doi.org/10.1080/10584609.2015.1102187>
- Tomé, J., & Cardita, S. (2021, décembre 20). In 2021, the Internet went for TikTok, space and beyond. *The Cloudflare Blog*. <http://blog.cloudflare.com/popular-domains-year-in-review-2021/>
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7), 1275-1289. <https://doi.org/10.3758/s13421-011-0104-1>
- Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological Inoculation against Misinformation: Current Evidence and Future Directions. *The ANNALS of the American Academy of Political and Social Science*, 700(1), 136-151. <https://doi.org/10.1177/00027162221087936>
- Treen, K. M. d'I., Williams, H. T. P., & O'Neill, S. J. (2020). Online misinformation about climate change. *WIREs Climate Change*, 11(5), e665. <https://doi.org/10.1002/wcc.665>
- Trippas, D., Pennycook, G., Verde, M. F., & Handley, S. J. (2015). Better but still biased: Analytic cognitive style and belief bias. *Thinking & Reasoning*, 21(4), 431-445. <https://doi.org/10.1080/13546783.2015.1016450>
- Tuchman, G. (2011, avril 26). *CNN investigation: Obama born in U.S.* <http://www.cnn.com/2011/POLITICS/04/25/birthers.obama.hawaii/index.html>
- Turnbull, M. (2017). *Speech introducing the National Security Legislation Amendment (Espionage and Foreign Interference)*.

- <https://www.malcolmturnbull.com.au/media/speech-introducing-the-national-security-legislation-amendment-espionage-an>
- United Airlines dément les rumeurs de faillite. (2008, septembre 8). *La Presse*.
<https://www.lapresse.ca/affaires/economie/200901/06/01-691548-united-airlines-dement-les-rumeurs-de-faillite.php>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 6(1), 2056305120903408. <https://doi.org/10.1177/2056305120903408>
- Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political Psychology in the Digital (mis)Information age: A Model of News Belief and Sharing. *Social Issues and Policy Review*, 15(1), 84-113.
<https://doi.org/10.1111/sipr.12077>
- Van Bavel, J. J., & Pereira, A. (2018). The Partisan Brain: An Identity-Based Model of Political Belief. *Trends in Cognitive Sciences*, 22(3), 213-224.
<https://doi.org/10.1016/j.tics.2018.01.004>
- Van Den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, 22(3), 335-351. <https://doi.org/10.1002/acp.1418>
- van der Linden, S. (2015). The conspiracy-effect: Exposure to conspiracy theories (about global warming) decreases pro-social behavior and science acceptance. *Personality and Individual Differences*, 87, 171-173. <https://doi.org/10.1016/j.paid.2015.07.045>
- van der Linden, S., Leiserowitz, A. A., Feinberg, G. D., & Maibach, E. W. (2014). How to communicate the scientific consensus on climate change: Plain facts, pie charts or metaphors? *Climatic Change*, 126(1-2), 255-262. <https://doi.org/10.1007/s10584-014-1190-4>
- van Huijstee, D., Vermeulen, I., Kerkhof, P., & Droog, E. (2021). Continued influence of misinformation in times of COVID-19. *International Journal of Psychology*, 57(1), 136-145. <https://doi.org/10.1002/ijop.12805>
- Vitriol, J., Sandor, J., Vidigal, R., & Farhart, C. (2022). On the Independent Roles of Cognitive & Political Sophistication: Variation Across Attitudinal Objects. *Applied Cognitive Psychology*, 37(2), 319-331. <https://doi.org/10.1002/acp.4022>
- Vlasceanu, M., & Coman, A. (2022). The impact of social norms on health-related belief update. *Applied Psychology: Health and Well-Being*, 14(2), 453-464.
<https://doi.org/10.1111/aphw.12313>
- Vohs, K. D., & Schooler, J. W. (2008). The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating. *Psychological Science*, 19(1), 49-54.
<https://doi.org/10.1111/j.1467-9280.2008.02045.x>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>
- Vraga, E., & Bode, L. (2017a). I do not believe you: How providing a source corrects health misperceptions across social media platforms. *Information, Communication & Society*, 21, 1-17. <https://doi.org/10.1080/1369118X.2017.1313883>

- Vraga, E., & Bode, L. (2017b). Using Expert Sources to Correct Health Misinformation in Social Media. *Science Communication*, 39(5), 621-645. <https://doi.org/10.1177/1075547017731776>
- Vraga, E., & Bode, L. (2020). Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation. *Political Communication*, 37(1), 136-144. <https://doi.org/10.1080/10584609.2020.1716500>
- Wagner-Egger, P. (2021). *Psychologie des croyances aux théories du complot : Le bruit de la conspiration* (Presses universitaires de Grenoble). <https://doi.org/10.3917/pug.wagne.2021.01>
- Wagner-Egger, P., & Gygax, P. (2018). Diana Was Not Involved in the 9/11 Terrorist Attacks! Or Was She? Newspaper Headlines and the Boomerang Effect. *Swiss Journal of Psychology*, 77(1), 15-22. <https://doi.org/10.1024/1421-0185/a000203>
- Walster, E., Berscheid, E., & Abrahams, D. (1967). Effectiveness of debriefing following deception experiments. *Journal of Personality and Social Psychology*, 6(4), 371-380. <https://doi.org/10.1037/h0024827>
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2019). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350-375. <https://doi.org/10.1080/10584609.2019.1668894>
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423-441. <https://doi.org/10.1080/03637751.2018.1467564>
- Walter, N., & Tukachinsky, R. (2020). A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It? *Communication Research*, 47(2), 155-177. <https://doi.org/10.1177/0093650219854600>
- Wardle, C. (2018). *Information Disorder: The Essential Glossary*. https://firstdraftnews.org/wp-content/uploads/2018/07/infoDisorder_glossary.pdf
- Warzel, C. (2018). *Believable: The Terrifying Future of Fake News*. BuzzFeed News. <https://www.buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news>
- Watson, C. A. (2018). Information Literacy in a Fake/False News World: An Overview of the Characteristics of Fake News and its Historical Development. *International Journal of Legal Information*, 46(2), 93-96. <https://doi.org/10.1017/jli.2018.25>
- Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology*, 92(5), 821-833. <https://doi.org/10.1037/0022-3514.92.5.821>
- Webster, D. M., & Kruglanski, A. W. (1994). Individual Differences in Need for Cognitive Closure. *Journal of Personality and Social Psychology*, 67(6), 1049-1062. <https://doi.org/10.1037/0022-3514.67.6.1049>
- Webster, F. (1999). Knowledgeability and democracy in an information age. *Library Review*, 48(8), 373-383. <https://doi.org/10.1108/00242539910288764>
- Wechsler, H., Nelson, T. E., Lee, J. E., Seibring, M., Lewis, C., & Keeling, R. P. (2003). Perception and reality: A national evaluation of social norms marketing interventions

- to reduce college students' heavy alcohol use. *Journal of Studies on Alcohol*, 64(4), 484-494. <https://doi.org/10.15288/jsa.2003.64.484>
- Weeks, B., & Garrett, R. K. (2014). Electoral Consequences of Political Rumors: Motivated Reasoning, Candidate Rumors, and Vote Choice during the 2008 U.S. Presidential Election. *International Journal of Public Opinion Research*, 26(4), 401-422. <https://doi.org/10.1093/ijpor/edu005>
- Wegener, D., & Petty, R. (1995). Flexible Correction Processes in Social Judgment: The Role of Naive Theories in Corrections for Perceived Bias. *Journal of personality and social psychology*, 68(1), 36-51. <https://doi.org/10.1037//0022-3514.68.1.36>
- Wegener, D., & Petty, R. (1997). The Flexible Correction Model: The Role of Naive Theories of Bias in Bias Correction. *Advances in Experimental Social Psychology*, 29, 141-208. [https://doi.org/10.1016/S0065-2601\(08\)60017-9](https://doi.org/10.1016/S0065-2601(08)60017-9)
- Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11). <https://doi.org/10.22215/timreview/1282>
- Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology Section A*, 40(2), 361-387. <https://doi.org/10.1080/02724988843000168>
- Williams, C. C., Kappen, M., Hassall, C. D., Wright, B., & Krigolson, O. E. (2019). Thinking theta and alpha: Mechanisms of intuitive and analytical reasoning. *NeuroImage*, 189, 574-580. <https://doi.org/10.1016/j.neuroimage.2019.01.048>
- Winneg, K., Kenski, K., & Jamieson, K. H. (2005). Detecting the Effects of Deceptive Presidential Advertisements in the Spring of 2004. *American Behavioral Scientist*, 49(1), 114-129. <https://doi.org/10.1177/0002764205279440>
- Wittenberg, C., Tappin, B. M., Berinsky, A. J., & Rand, D. G. (2021). The (minimal) persuasive advantage of political video over text. *Proceedings of the National Academy of Sciences*, 118(47), e2114388118. <https://doi.org/10.1073/pnas.2114388118>
- Wixted, J. T., & Ebbesen, E. B. (1991). On the Form of Forgetting. *Psychological Science*, 2(6), 409-415. <https://doi.org/10.1111/j.1467-9280.1991.tb00175.x>
- Wolburg, J. M. (2006). College Students' Responses to Antismoking Messages: Denial, Defiance, and Other Boomerang Effects. *Journal of Consumer Affairs*, 40(2), 294-323. <https://doi.org/10.1111/j.1745-6606.2006.00059.x>
- Wolf, S., & Montgomery, D. A. (1977). Effects of Inadmissible Evidence and Level of Judicial Admonishment to Disregard on the Judgments of Mock Jurors¹. *Journal of Applied Social Psychology*, 7(3), 205-219. <https://doi.org/10.1111/j.1559-1816.1977.tb00746.x>
- Wood, T., & Porter, E. (2016). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 41(1), 135-163. <https://doi.org/10.1007/s11109-018-9443-y>
- Xu, Y., Wong, R., He, S., Veldre, A., & Andrews, S. (2020). Is it smart to read on your phone? The impact of reading format and culture on the continued influence of misinformation. *Memory & Cognition*, 48(7), 1112-1127. <https://doi.org/10.3758/s13421-020-01046-0>

- Yadav, A., Phillips, M. M., Lundeberg, M. A., Koehler, M. J., Hilden, K., & Dirkin, K. H. (2011). If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *Journal of Computing in Higher Education*, 23(1), 15-37. <https://doi.org/10.1007/s12528-011-9042-y>
- Young, D. G., Jamieson, K. H., Poulsen, S., & Goldring, A. (2018). Fact-Checking Effectiveness as a Function of Format and Tone: Evaluating FactCheck.org and FlackCheck.org. *Journalism & Mass Communication Quarterly*, 95(1), 49-75. <https://doi.org/10.1177/1077699017710453>
- Zaccoletti, S., Raccanello, D., Burro, R., & Mason, L. (2023). Reading with induced worry: The role of physiological self-regulation and working memory updating in text comprehension. *British Journal of Educational Psychology*, 93(1), 26-47. <https://doi.org/10.1111/bjep.12491>
- Zimmermann, F., & Kohring, M. (2020). Mistrust, Disinforming News, and Vote Choice: A Panel Survey on the Origins and Consequences of Believing Disinformation in the 2017 German Parliamentary Election. *Political Communication*, 37(2), 215-237. <https://doi.org/10.1080/10584609.2019.1686095>
- Zollo, F., Bessi, A., Vicario, M. D., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S., & Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLOS ONE*, 12(7), e0181821. <https://doi.org/10.1371/journal.pone.0181821>
- Zukier, H., & Pepitone, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology*, 47, 349-360. <https://doi.org/10.1037/0022-3514.47.2.349>
- Zylberberg, S. (2018, octobre 29). *L'énigme du nénuphar et de l'étang*. <https://jeretiens.net/enigme-nenuphar-etang-combien-jours/>

Titre : L'influence persistante des fausses informations malgré leur correction : étude de validité et de nouveaux déterminants

Mots-clés : fausse information – effet d'influence continue – correction – fact-checking

Résumé : De nombreux travaux sur l'« effet d'influence continue » (Johnson & Seifert, 1994) observent qu'une correction est parfois insuffisante voire inefficace pour lutter contre l'adhésion à une fausse information. A une époque où les fausses informations constituent une problématique sociétale majeure et qu'une stratégie fréquemment utilisée pour lutter contre leur influence consiste à les corriger, il apparaît nécessaire de mieux comprendre cet effet. Cette thèse présente une synthèse inédite des explications théoriques de l'effet d'influence continue et présente cinq recherches basées sur des méthodes expérimentale ou corrélacionnelle qui examinent la validité et les déterminants de cet effet. Les résultats confirment sa validité tout en mettant en évidence que d'autres effets restent possibles à la suite de la correction d'une fausse information, tels qu'une correction efficace ou un effet de sur-correction. Ces travaux ont aussi permis d'observer que de bonnes performances en mémoire de travail réduiraient cet effet. Cependant, aucun effet protecteur du névrosisme et de l'implication dispositionnelle ou situationnelle n'a été relevé. Ces différentes observations invitent à poursuivre la recherche autour des variables capables de réduire l'effet d'influence continue.

Title: The persistent influence of false information despite correction: a study of validity and new determinants

Keywords: misinformation – continued influence effect – correction – fact-checking

Abstract: Numerous studies on the “continued influence effect” (Johnson & Seifert, 1994) have observed that corrections are sometimes insufficient or even ineffective in reducing belief in misinformation. In an era where misinformation poses a significant societal challenge, and correcting it is a widely used strategy, it has become essential to gain a deeper understanding of this effect. This thesis presents an original synthesis of the theoretical explanations of the continued influence effect and reports five studies using experimental or correlational methods that examine the validity and determinants of this effect. The results confirm its validity while also highlighting that alternative outcomes remain possible, such as an effective correction or even an over-correction effect. These studies also found that strong working memory performance may reduce this effect. However, no protective role of neuroticism or dispositional and situational involvement was identified. These findings underscore the need for further research into variables that can reduce the continued influence effect.

