



THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 2 - Jean Jaurès

Présentée et soutenue par
Valentin GOULETTE

Le 15 septembre 2023

La rationalité des jugements causaux : L'effet de la valence morale à l'épreuve des liens causaux et de la logique conversationnelle

Ecole doctorale : **CLESCO - Comportement, Langage, Education, Socialisation, Cognition**

Spécialité : **Psychologie**

Unité de recherche :

CLLE - Unité Cognition, Langues, Langage, Ergonomie

Thèse dirigée par
Fanny VERKAMPT

Jury

Mme Valérie PENNEQUIN, Rapporteur
M. Pascal WAGNER-EGGER, Rapporteur
M. Alain SOMAT, Examineur
M. Bastien TREMOLIERE, Examineur
Mme Isabelle MILHABET, Examinatrice
Mme Fanny VERKAMPT, Directrice de thèse

À mon mentor, Denis Hilton

Résumé

Cette thèse s'intéresse à l'effet de la valence morale sur le jugement causal. L'effet en question survient lorsqu'un agent possédant une valence morale négative (vs. neutre ou positive) est davantage désigné comme la cause d'un dommage. Le Modèle du Contrôle Coupable soutient que le jugement causal est intrinsèquement biaisé par les réactions évaluatives des observateurs sociaux envers l'agent ayant une valence morale négative. Cependant, ni ce modèle ni le reste de la littérature ne précise l'ampleur de ce biais de jugement. Par exemple, la question de savoir si les individus iraient jusqu'à inventer un lien causal inexistant entre l'agent et le dommage reste en suspens. Ce biais de jugement causal étant à craindre dans certains domaines, notamment judiciaire, il semble nécessaire d'étudier sa nature et les conditions dans lesquelles il se manifeste. Par conséquent, l'objectif principal de cette thèse était d'examiner dans quelle mesure le jugement causal de sens commun est biaisé par les évaluations négatives suscitées par la valence morale négative d'un agent.

Pour atteindre cet objectif, une première série d'études ($n = 1081$) a été menée afin de circonscrire l'effet de la valence morale et d'étudier sa relation avec deux critères normatifs fondamentaux de la causalité, à savoir la présence d'un lien causal et la force de ce lien. Les participants ont été invités à lire un récit dans lequel un agent, dont la valence morale était manipulée, et un dommage étaient présentés. Le lien causal entre l'agent et le dommage a également été manipulé (e.g., fort, faible, absent). Après la lecture du récit, les participants étaient invités à répondre à plusieurs questions, mesurant notamment le jugement causal. Les résultats obtenus semblent indiquer que la recherche d'informations peut être biaisée par l'effet de la valence morale, cela uniquement dans des situations où il existe un lien causal entre les actions de l'agent et le dommage. De plus, il semble que la force du lien causal soit également prise en compte par les individus. La grande majorité de nos participants désignait la condition ayant la plus grande force causale comme cause principale du dommage, alors même qu'un agent possédant une valence morale négative était impliqué dans l'événement. Néanmoins, l'effet du type de question posée (ouverte vs. centrée sur l'agent) sur la réplication de l'effet de la valence morale s'est avéré inattendu. Cette observation nous a conduits à examiner le rôle de l'interprétation des questions causales au sein de la tâche de jugement, en raison de la polysémie du terme *cause*. En nous appuyant sur l'approche pragmatique du jugement causal, nous avons analysé les récits utilisés dans la littérature pour mettre en évidence un problème méthodologique dans les études examinant l'effet de la valence morale. Deux études supplémentaires ($n = 407$) ont ensuite été menées afin de déterminer si le cadrage conversationnel (i.e., les informations guidant l'interprétation de la question causale) neutraliserait l'effet de la valence morale. Nos données n'ont pas corroboré cette hypothèse, renforçant alors l'hypothèse d'un biais de jugement plutôt que celle de l'ambivalence du mot *cause*.

L'ensemble des résultats de cette thèse apporte de nouvelles perspectives, tant sur le plan théorique qu'appliqué, quant à la rationalité des jugements causaux de sens commun. D'une part, les individus respectent prioritairement les deux critères normatifs fondamentaux de la causalité. D'autre part, l'effet de la valence morale ne semble pas s'expliquer par le cadrage de la question de jugement causal. En somme, le travail réalisé permet d'établir l'ampleur du biais de jugement que constitue l'effet de la valence morale, tout en offrant des orientations pour de futures études visant à tester des méthodes de débiaisement.

Mots clés : Jugement causal, rationalité, biais, pragmatique, blâme

Abstract

This thesis focuses on the effect of moral valence on causal judgment. The effect in question occurs when an agent with a negative (vs. neutral or positive) moral valence is more likely to be designated as the cause of a harm. The Culpable Control Model argues that causal judgment is intrinsically biased by the evaluative reactions of social observers towards the agent with a negative moral valence. However, neither this model nor the rest of the literature specifies the extent of this judgment bias. For example, the question of whether individuals would go so far as to invent a non-existent causal link between the agent and the harm remains unanswered. As this causal judgement bias is to be feared in certain fields, notably judicial, it seems necessary to study its nature and the conditions under which it manifests itself. Consequently, the main aim of this thesis was to examine the extent to which common-sense causal judgment is biased by negative evaluations induced by an agent's negative moral valence.

To achieve this aim, a first series of studies ($n = 1081$) was conducted to circumscribe the effect of moral valence and investigate its relationship with two fundamental normative criteria of causality, namely the presence of a causal link and the strength of that link. Participants were asked to read a story in which an agent, whose moral valence was manipulated, and a harm were presented. The causal link between the agent and the damage was also manipulated (e.g., strong, weak, absent). After reading the story, participants were asked to answer several questions, including causal judgment questions. The results suggest that information seeking may be biased by the effect of moral valence, but only in situations where there is a causal link between the agent's actions and the harm. Moreover, it seems that the strength of the causal link is also considered by individuals. Most of our participants designated the condition with the greatest causal strength as the main cause of the harm, even though an agent with a negative moral valence was involved in the event. Nevertheless, the effect of the type of question asked (open-ended vs. agent-centered) on the replication of the moral valence effect was unexpected. This observation led us to examine the role of interpretation in causal questions within the judgment task, given the polysemy of the term *cause*. Drawing on the pragmatic approach to causal judgment, we examined the narratives used in the literature to highlight a methodological problem in studies investigating the effect of moral valence. Two additional studies ($n = 407$) were then conducted to determine whether conversational framing (i.e., information guiding the interpretation of the causal question) would neutralize the effect of moral valence. However, our data did not corroborate this hypothesis, further supporting the judgment bias hypothesis rather than that of the ambivalence of the word *cause*.

The overall results of this thesis provide new insights, both theoretical and applied, into the rationality of common-sense causal judgments. On the one hand, individuals give priority to the two fundamental normative criteria of causality. On the other hand, the effect of moral valence does not seem to be explained by the framing of the causal judgment question. In sum, this work establishes the magnitude of the judgment bias constituted by the effect of moral valence, while offering guidelines for future studies aimed at testing debiasing methods.

Key words: Causal judgment, rationality, bias, pragmatics, blame

Remerciements

Je souhaite exprimer ma gratitude envers toutes les personnes qui, de manière directe ou indirecte, ont joué un rôle dans la réalisation de cette thèse qui, je l'espère, sera jugée apte à être soutenue.

J'éprouve un profond sentiment de gratitude envers les deux personnes qui ont encadré ce travail de thèse. Je suis extrêmement reconnaissant envers Denis Hilton, qui m'a initié à la recherche et m'a partagé sa passion pour l'étude du jugement causal dès ma première année de Master. Si je suis arrivé jusqu'ici, c'est en grande partie grâce à vous. Je n'oublierai jamais tout ce que vous avez fait pour moi.

Je souhaite également remercier Fanny Verkampt pour m'avoir encadré avec autant d'implication et de bienveillance. Merci pour la précision de tes retours, ta disponibilité et tes encouragements qui m'ont permis de m'améliorer et de me sentir soutenu dans les moments difficiles.

J'adresse mes remerciements aux membres de mon jury. Je suis convaincu que leurs retours me permettront de prendre du recul sur le travail effectué et d'en appréhender la portée ainsi que les limites. Merci aux rapporteurs, Valérie Pennequin et Pascal Wagner-Egger, d'avoir accepté cette charge de travail supplémentaire. Merci à Alain Somat pour les précieux conseils prodigués lors des comités de suivi de thèse et pour avoir accepté d'examiner la version finale de ce travail. Je remercie également Bastien Trémolière et Isabelle Milhabet d'avoir généreusement accepté d'endosser le rôle d'examineurs.

Je suis reconnaissant envers Valérie Le Floch et Patrice Terrier pour leurs retours très pertinents sur mon travail ainsi que pour leurs conseils avisés.

Je tiens à exprimer ma profonde gratitude à Dimitri, Quentin, Maxime et Jérémy pour les relectures, mais aussi pour leur précieuse amitié. Toutes ces conversations passionnantes avec vous ont joué un rôle essentiel dans le développement de ma conception de la psychologie et de la recherche scientifique.

Merci à Julie Vincent pour son expertise juridique. Je lui suis reconnaissant d'avoir généreusement consacré de son temps pour m'éviter de dire trop de bêtises sur le sujet.

Merci à tous les collègues, partenaires de RU et amis doctorants et post-doctorants du CLLE pour tous ces bons moments passés ensemble : Jean, Colin, Océane, Daniele, Mariame,

Robin, Florian, Efisio, Noé et tous les autres que j'ai pu oublier. Votre présence, votre soutien et tous les rires que nous avons partagé ont contribué à rendre cette expérience mémorable.

Je remercie également tous les membres passés et présents d'InCOGnu, qui sont devenus pour beaucoup des amis. Je vous remercie pour l'ambiance fabuleuse qui règne au sein de l'association, cette amour pour les sciences cognitives et cette volonté de partager les connaissances scientifiques au plus grand nombre. Je suis honoré de faire partie de cette famille scientifique.

Merci à mes collègues représentantes des doctorants CLESCO : Marie-Amélie, Lucile, Chloé, Anne-Flore et Nathalie. Ce mandat n'aurait pas été aussi agréable sans vous.

Merci à mes amis Elohan et Valentin pour tous les bons moments que nous avons partagé et pour avoir su me rappeler de ne pas se laisser absorber entièrement par la thèse.

Evidemment, je suis infiniment reconnaissant envers mes proches et ma famille pour être là dans les bons, comme dans les mauvais moments. Merci à ma famille de me soutenir dans tout ce que j'entreprends et de m'avoir donné très tôt le goût de la connaissance. Votre amour a été une source de motivation constante.

Merci également à ma partenaire Nolwenn, dont le soutien a été particulièrement décisif, surtout dans les derniers moments avant le rendu du manuscrit. Ta présence a été et continue d'être inestimable. Je suis reconnaissant de t'avoir à mes côtés dans cette aventure.

Enfin, je remercie tous les participants ayant accepté de prendre de leur temps pour participer aux études de cette thèse sans qui, cette dernière n'aurait pas pu voir le jour.

Table des matières

Introduction générale	1
De l'attribution causale au jugement causal	2
La rationalité des jugements	4
Objectifs de la thèse	7
Perspective d'ensemble du travail de thèse	8
Paradigme expérimental et précautions scientifiques	10
Chapitre 1. Les critères normatifs du jugement causal	13
D'un point de vue philosophique et logique	14
Les deux approches philosophiques de la causalité	14
De la nécessité à la nécessité réelle	16
D'une causalité dichotomique à une causalité graduée	17
Le modèle causal d'une situation	19
D'un point de vue juridique	22
La présomption de rationalité	22
Déterminer un lien causal entre les actions de l'accusé et le dommage	24
Conclusion	29
Chapitre 2. De l'effet des normes à la validation du blâme	30
Effet de l'infraction d'une norme morale	31
Le courant de l'Homme comme avocat	33
La philosophie légale de Hart et Honoré	34
Un jugement séquentiel et rationnel	36
Limites du courant de « l'Homme comme avocat »	37
L'approche du raisonnement contrefactuel	39
Déterminer la nécessité réelle	40
Déterminer la sensibilité	43
Un jugement causal rationnel ?	45
La focalisation sur la condition anormale	46
Limites de l'approche contrefactuelle	50
Conclusion	53
La dimension motivée du jugement causal	54
L'évaluation des états mentaux de nos semblables	55
Les évaluations négatives spontanées	58
Le Modèle du Contrôle Coupable et le jugement causal	59

Distinction entre raisonnement contrefactuel et CCM.....	62
Conclusion	64
Chapitre 3. Délimiter l'effet de la valence morale sur le jugement causal	65
Processus de validation du blâme et nécessité réelle	67
Processus de validation du blâme et sensibilité	71
Études 1 et 2 : Effet de la valence morale sur le jugement causal en fonction des critères normatifs de nécessité réelle et de sensibilité	74
Étude 1 : Effet de la valence morale sur le jugement causal en fonction des critères normatifs de nécessité réelle et de sensibilité avec une question ouverte.....	74
Méthode	76
Résultats.....	80
Discussion de l'Étude 1	83
Étude 2 : Effet de la valence morale sur le jugement causal en fonction des critères normatifs de nécessité réelle et de sensibilité avec une question à échelle de réponse	87
Méthode	87
Résultats.....	88
Discussion de l'Étude 2	91
Discussion générale des Études 1 et 2	92
Étude 3 : Effet de la valence morale sur le jugement causal en fonction du critère normatif de nécessité réelle avec un agent aux motivations réellement louables....	95
Méthode	97
Résultats.....	100
Discussion de l'Étude 3	109
Étude 4 : Effet de la valence morale sur le jugement causal en fonction du critère normatif de nécessité réelle et de la sévérité du dommage avec un agent aux motivations réellement louables	113
Méthode	114
Résultats.....	115
Discussion de l'Étude 4	121
Discussion du Chapitre 3	124
Chapitre 4. Une perspective pragmatique de l'effet de la valence morale	130
La logique conversationnelle et les biais de raisonnement.....	131
L'approche pragmatique du jugement causal	134
Les implications pragmatiques des scénarios et de la focalisation de la question.	139
Revue narrative de la littérature.....	141
Analyse pragmatique des modèles causaux et de la focalisation de la question.....	143

Effet du cadrage conversationnel de la tâche sur le jugement causal	149
Étude 5 : Influence du cadrage conversationnel sur l'effet de la valence morale de l'agent sur le jugement causal.....	151
Méthode	151
Résultats.....	154
Discussion de l'étude 5	158
Étude 6 : Influence du cadrage conversationnel sur l'effet de la valence morale de l'agent sur le jugement causal (nouvelle opérationnalisation du cadrage)	162
Méthode	162
Résultats.....	164
Discussion de l'étude 6	168
Discussion du Chapitre 4	169
Discussion générale	175
Apports de la thèse.....	176
Connaissance et application prioritaire des critères normatifs de la causalité... ..	177
Interaction entre l'application des critères normatifs et l'effet de la valence morale	178
Effet de la valence morale : biais de jugement ou logique conversationnelle ?	181
Des différences inter-individuelles dans la rationalité des jugements causaux ?.....	182
Résumé des apports de la thèse.....	184
Limites et perspectives.....	186
La validité écologique des études sur l'effet de la valence morale.....	187
Plusieurs définitions de la rationalité.....	192
De la mesure des évaluations spontanées à la généralisation des résultats.....	194
Conclusion	198
Références	200

Introduction générale

« Le sens commun se nourrit ainsi de causalité et il serait entièrement désorienté dans un monde sans causes où tout serait surprise et événement. » (Malherbe, 1994, p.6). Ainsi parlait le philosophe Michel Malherbe de ce qui semble être le besoin fondamental de l'être humain de comprendre son environnement physique et social (Waldmann, 2017). Un fait notable est que la plupart des investigations causales sont enclenchées par des événements négatifs et inattendus (Hastie, 1984 ; Weiner, 1985 ; Wong & Weiner, 1981), qu'il s'agisse d'événements publics, tels que l'explosion d'une navette spatiale, un attentat terroriste, une catastrophe naturelle, ou d'événements privés tels qu'un accident ou un divorce. Ceci est d'autant plus vrai lorsque ces événements négatifs et inattendus ont été provoqués par des comportements humains volontaires (Hart & Honoré, 1985 ; Heider, 1958). Le présent travail de thèse se focalise précisément sur le jugement causal dans les situations à l'issue négative et impliquant un agent humain (e.g., meurtre, vol, etc.).

Malgré la nécessité de détecter les relations causales face à de tels événements, celles-ci ne sont pas directement perceptibles, et nous devons donc les inférer (Buehner, 2017 ; White, 1990). Par conséquent, comme toute opération de raisonnement, les individus peuvent commettre des erreurs, notamment en raison de biais cognitifs (e.g., Tversky & Kahneman, 1974 ; Wason, 1960). Selon le Modèle du Contrôle Coupable (CCM, Alicke 2000 ; Alicke et al., 2008, 2011), le jugement causal est intrinsèquement biaisé dans les situations qui nous intéressent. Selon ce modèle, lorsque l'agent humain impliqué dans l'événement négatif suscite une évaluation négative (e.g., en raison de la valence morale de ses états mentaux) de la part de l'observateur social, un *processus de validation du blâme* inciterait ce dernier à surestimer le lien causal entre les actions de l'agent et le dommage. Cette hypothèse a été corroborée par

les nombreuses répliques d'un effet : pour une même action, un agent humain est davantage désigné comme cause d'un dommage lorsque ses états mentaux, sa personnalité ou ses actions possèdent une valence morale négative plutôt que neutre ou positive (e.g., Alicke, 1992 ; Alicke et al., 2008, 2011 ; Lagnado & Channon, 2008 ; Phillips & Shaw, 2015 ; Rogers et al., 2019). Il s'agirait alors d'un biais cognitif remettant en question la rationalité des jugements causaux de sens commun. Bien que le jugement causal soit une composante essentielle de la cognition humaine, le domaine judiciaire est particulièrement concerné. En effet, pour désigner un individu coupable d'un dommage et le sanctionner en conséquence, la première étape pour les juges et jurés est d'établir un lien causal entre ses actions et le dommage (Hart & Honoré, 1985 ; Summers, 2018). Par conséquent, l'étude de la rationalité des jugements causaux de sens commun est particulièrement cruciale pour ce domaine.

Avant d'aller plus loin, nous allons apporter certaines précisions sur les termes employés dans ce travail de thèse. En premier lieu, en raison des nombreuses facettes de l'étude du *jugement causal* et de l'ambiguïté du terme, nous préciserons sur quel type de jugement causal portera ce travail. En deuxième lieu, nous définirons ce que nous entendons par *biais de jugement* et *rationalité des jugements*. Une fois ces définitions établies, nous présenterons les objectifs et les questions de recherche auxquels cette thèse vise à apporter des réponses. La perspective d'ensemble du travail de thèse sera ensuite exposée. Enfin, nous présenterons le paradigme expérimental utilisé pour parvenir à nos objectifs de recherche, ainsi que certaines précautions méthodologiques et statistiques adoptées visant à améliorer la validité et la reproductibilité de nos résultats.

De l'attribution causale au jugement causal

Dès les années 50, la recherche en psychologie s'est intéressée au jugement causal (pour un historique, voir Alicke et al., 2015). Selon Hilton (2007), il existe une confusion importante

dans la littérature à propos de la question de la rationalité des jugements causaux de sens commun. Une raison à cela est que cette faculté a reçu des appellations très variées et rarement distinguées telles que l'attribution causale, l'explication causale, le jugement causal ou encore le raisonnement causal. Or, les processus d'inférence causale, les critères pertinents de rationalité et les domaines d'application auxquels renvoient ces différentes appellations ne sont pas nécessairement identiques (e.g., Buss, 1978 ; Fincham & Jaspars, 1980 ; Hamilton, 1980). Par conséquent, afin d'éviter toute ambiguïté, il nous paraît notamment nécessaire de démarquer notre domaine d'étude de celui de l'attribution causale, couramment étudié.

La psychologie sociale s'est considérablement intéressée au jugement causal sous le terme d'*attribution causale* avec les travaux pionniers de Fritz Heider (1944, 1946, 1958). Cette perspective de recherche très prolifique portait sur l'étude de la compréhension du monde social. Son objectif était d'étudier la façon dont les individus expliquent le comportement d'autrui, et, dans une moindre mesure, leur propre comportement (e.g., Fischer, 2020 ; Malle, 2011 ; Moscovici, 1972). Les principales théories de l'attribution causale (Heider, 1958 ; Jones & Davis, 1965 ; Kelley, 1967, 1973) supposent que, pour expliquer un comportement, les individus font majoritairement appel à deux types d'explication : des explications internes à l'individu qui a émis le comportement (e.g., effort, motivation, personnalité) et des explications qui lui sont externes (e.g., l'environnement, difficulté de la tâche). Selon cette perspective, l'objectif principal de l'observateur social est d'identifier des propriétés stables à autrui sur la base d'observations de comportements fluctuants (Heider, 1958, p.79). En somme, ces théories sont très spécifiques à la recherche des causes des comportements humains. Or, nos travaux se concentrent sur l'étude des explications de dommages (e.g., meurtres, blessures corporelles, dégâts matériels), plutôt que sur les explications de comportements humains ayant pu les provoquer. Dans nos études, les raisons du comportement de l'agent humain seront systématiquement mentionnées aux participants. Par conséquent, afin d'éviter toute ambiguïté

théorique, nous n'utiliserons volontairement pas le terme *attribution causale*, renvoyant davantage à l'explication d'un comportement humain, notamment en termes personnologiques ou environnementaux. Dans ce travail de thèse, nous parlerons de *jugement causal*¹ pour désigner l'évaluation rétrospective des causes d'un dommage particulier.

Le terme *jugement causal* étant à présent défini et distingué de celui d'*attribution causale*, nous allons définir ce que nous entendons par *biais de jugement* et mettre en évidence l'importance de l'étude de la rationalité des jugements de sens commun.

La rationalité des jugements

Un *biais de jugement* est un écart systématique entre les jugements de sens commun et un point de référence considéré comme rationnel (e.g., Caverni et al., 1990 ; Korteling & Toet, 2020). Ce point de référence est ce que l'on nomme généralement un modèle normatif. Le terme *normatif* permet de décrire « l'utilisation d'une règle lorsqu'il existe un consensus parmi les scientifiques du domaine sur le fait que la règle est appropriée pour le problème particulier » (Nisbett & Ross, 1980, p. 13). Une fois ce point de référence défini, la rationalité des individus est évaluée à partir de la distance entre les comportements observés et la réponse fournie par le modèle normatif (Etzioni, 2014). La notion de *biais* s'inscrit alors nécessairement dans une définition normative de la rationalité. Néanmoins, cette définition de la rationalité n'est pas la seule (Evans, 2007) et a suscité de nombreux débats quant à son utilisation dans la recherche en psychologie (pour une revue en français, Wagner-Egger, 2011). Il s'agit pourtant de la définition la plus utilisée en sciences cognitives car celle-ci présente un certain nombre d'avantages (Stanovich et al., 2016).

¹ Nous utiliserons le terme « jugement causal » plutôt que « raisonnement causal », même si la distinction entre les deux termes n'est pas clairement établie dans la littérature. Le terme « raisonnement causal » semble néanmoins plus général que ce qui nous intéresse (Bowers, 2021 ; Waldmann, 2017) et peut renvoyer à l'inférence de liens causaux généraux ou au raisonnement prospectif (pour une brève discussion sur la distinction entre ces termes, voir Operskalski & Barbey, 2017, p.28).

L'une des principales justifications de l'utilisation d'une telle définition de la rationalité dans la recherche en psychologie est que les biais (i.e., écarts par rapport au modèle normatif) ont des effets indésirables sur la société (Caverni et al., 1990 ; Toplak et al., 2017). Par exemple, il ne faudrait pas hésiter à qualifier d'irrationnels l'effet de l'origine ethnique (Anwar et al., 2012 ; Mitchell, 2005) ou de l'attractivité physique de l'accusé (e.g., Efran, 1974 ; Saladin et al., 1988) sur les décisions judiciaires, ou encore certains raisonnements dans le milieu médical qui augmenteraient les erreurs de diagnostics (Ludolph & Shultz, 2018). Comme le suggèrent Reyna et ses collaborateurs (2003), il ne faut donc pas hésiter à étiqueter clairement un comportement comme irrationnel afin que la recherche sur les jugements humains soit pertinente pour les problèmes du monde réel. Au-delà de l'importance de la détection des biais et de l'étude de leurs conditions d'apparition, la question de leur réduction est également cruciale. L'utilisation d'un modèle normatif permet alors de fixer un objectif idéal vers lequel les psychologues peuvent tenter d'amener les jugements biaisés, grâce aux méthodes de débiaisement notamment (e.g., Morewedge et al., 2015 ; Sellier et al., 2019 ; Symborski et al., 2014). La réduction des biais permettrait ainsi d'améliorer l'efficacité des jugements de sens commun et de réduire les conséquences sociétales négatives qui en découlent.

Par ailleurs, la question de la rationalité des jugements de sens commun implique des enjeux démocratiques interrogeant la place du citoyen dans les décisions sociétales (Stiegler, 2019). Comme le soulignait Evans (1994), l'un des pionniers de la recherche sur les biais de jugement, si nous détectons un biais dans une tâche donnée, par exemple une décision judiciaire ou médicale, deux solutions s'offrent à nous : réduire les biais de jugement, ou remplacer les individus par des systèmes informatiques plus proches du modèle normatif pertinent pour la tâche. Dans le domaine judiciaire, cette deuxième solution a déjà été appliquée dans certains pays. Nous pouvons notamment citer l'Ontario (Canada) où une intelligence artificielle (IA)

est chargée de trancher les conflits entre voisins ou entre salarié et employeur, ou encore l'Estonie, où une IA juge la culpabilité d'un accusé pour des litiges mineurs (inférieurs à 7000 euros)². En France, une première tentative d'intégration de l'IA dans l'aide aux décisions judiciaires a été effectuée par le décret n° 2020-356 du 27 mars 2020 sur la mise en place d'un traitement automatisé des décisions sur les préjudices corporels dénommé *DataJust*. Néanmoins, le développement de cet algorithme a été abandonné, notamment pour des raisons de complexité de la notion de préjudice corporel³. En somme, l'étude des biais de jugement et de leur réduction est un enjeu crucial quant à la place de l'individu dans les décisions de la vie courante. De plus, ces recherches permettent également d'améliorer la qualité des décisions judiciaires, médicales et même scientifiques (e.g., Dror, 2018), sans pour autant supplanter l'individu par un système informatique.

Bien que l'étude de la rationalité des jugements soit primordiale tant pour les effets indésirables des biais que pour la place de l'individu dans la société, un aspect a souvent été négligé (De Neys & Bonnefon, 2013 ; Stanovich, 2018). Il s'agit de l'identification de la nature du biais de jugement étudié. En effet, la rationalité a souvent été conçue de façon dichotomique : un jugement est considéré comme biaisé ou non. Depuis une cinquantaine d'années, la présence d'un biais a souvent été essentiellement expliquée par l'influence d'une heuristique inappropriée, et par l'absence d'activation d'un raisonnement délibéré et coûteux en ressources cognitives (pour une revue des théories Dual Process, De Neys, 2018). Néanmoins, cette conception a récemment été discutée et développée (Bago & De Neys, 2019 ; De Neys & Pennycook, 2019 ; Evans, 2019). Un jugement pourrait s'écarter du modèle normatif pour plusieurs raisons et certains biais seraient plus « irrationnels » que d'autres

² Informations retrouvées sur le site du gouvernement français : <https://www.vie-publique.fr/eclairage/277098-lintelligence-artificielle-ia-dans-les-decisions-de-justice>

³ <https://acteurspublics.fr/articles/exclusif-le-ministere-de-la-justice-renonce-a-son-algorithme-datajust>

(Brainerd & Reyna, 2001). En somme, il y aurait des biais de nature différente. Dans cette perspective, certains auteurs ont proposé des taxonomies de biais de jugement (e.g., Brainerd & Reyna, 2001 ; Stanovich, 2018 ; Stanovich et al., 2016 ; Reyna et al., 2003). Par exemple, selon Stanovich (2018), un biais peut être dû à une absence de connaissance des critères normatifs appropriés pour la tâche, une incapacité à détecter que la réponse fournie est inappropriée, ou encore une incapacité à l'inhiber. Identifier la nature du biais étudié permettrait alors de déterminer l'intervention la plus pertinente pour le réduire ou le neutraliser. Si le biais observé est lié à une absence de connaissance des critères normatifs pertinents pour la tâche, il convient de former les individus à ces critères. En revanche, s'il s'agit d'un problème de détection ou d'inhibition, des méthodes d'inhibition de la réponse inappropriée (Houdé et al., 2000 ; Moutier et al., 2002) ou de détection de l'interférence entre deux processus (De Neys, 2014) seraient alors recommandées. C'est à travers cette conception « graduée » de la rationalité que nous avons mené nos travaux sur le jugement causal.

Objectifs de la thèse

Nous avons vu que l'effet de la valence morale d'un agent humain sur le jugement causal a été répliqué de nombreuses fois. Le CCM (Modèle du Contrôle Coupable) postule que cet effet est la preuve que le jugement causal de sens commun est intrinsèquement biaisé. Néanmoins, ce modèle, comme le reste de la littérature, ne précise pas l'amplitude de l'écart entre les jugements causaux et le modèle normatif pertinent pour la tâche. Par exemple, nous pouvons nous demander si les individus iraient jusqu'à inventer un lien causal inexistant entre l'agent et le dommage. De manière plus générale, la question de recherche principale qui nous intéressera est la suivante : Dans quelle mesure le jugement causal de sens commun est-il biaisé dans les situations à l'issue négative et impliquant un agent humain ? Pour répondre à cette question, il est nécessaire d'étudier l'interaction entre les critères normatifs et extra-normatifs dans le jugement causal des individus. D'une part, la réponse à cette question de

recherche permettra d'enrichir le cadre théorique en précisant la nature et les conditions d'apparition de ce biais de jugement. D'autre part, ces travaux permettront de guider les méthodes pour neutraliser ou réduire ce biais, notamment dans les décisions judiciaires.

Perspective d'ensemble du travail de thèse

Nous avons vu précédemment que pour être en mesure de qualifier si un jugement est biaisé ou non, il convient de définir dans un premier temps ce qui constitue un « bon » jugement causal. Par ailleurs, si nous voulons déterminer dans quelle mesure les individus s'écartent des critères normatifs, il est évidemment nécessaire de définir ces derniers de manière précise. Par conséquent, le Chapitre 1 aura pour objectif de définir le modèle normatif du jugement causal. Nous verrons que ce modèle repose sur deux critères utilisés tant par les philosophes, les juristes que par les psychologues sous des appellations différentes. Ces deux critères renvoient respectivement à la présence et à la force d'un lien causal. Nous verrons comment chaque discipline se les approprie.

Une fois ce cadre posé, il servira de base pour mener notre analyse sur la rationalité des jugements causaux de sens commun. Dans le Chapitre 2, nous aborderons le versant psychologique du jugement causal en présentant un effet très largement répliqué en psychologie expérimentale : la sélection de la condition anormale comme cause d'un dommage. Toujours dans ce chapitre, nous passerons en revue les modèles psychologiques permettant d'expliquer ce phénomène. Dans un premier temps, seront présentés les modèles rationalistes considérant que les individus sont doués pour évaluer les critères normatifs décrits dans le Chapitre 1, et pour comprendre leur environnement. Nous verrons toutefois que le respect de ces critères a principalement été étudié vis-à-vis d'événements physiques (e.g., boules de billards). Or, nous présenterons également certains travaux soulignant l'importance de dissocier les jugements vis-à-vis d'objets physiques (monde physique) de ceux émis vis-à-

vis d'agents humains (monde social). A travers les jugements formulés vis-à-vis du monde physique, l'objectif des individus serait de comprendre les événements qui les entourent pour s'y adapter efficacement (e.g., ne pas mettre de verre d'eau dehors, sinon il gèle). Ceux formulés à l'égard du monde social viserait un autre objectif, celui d'adapter nos comportements en tenant compte de notre environnement social. Cela implique de savoir évaluer nos semblables, et notamment leurs états mentaux pour pouvoir identifier les « méchants » (à éviter et à blâmer) et les « gentils » (à fréquenter et à louer). Par conséquent, nous verrons dans un second temps que pour expliquer un événement du monde social, d'autres variables s'ajouteraient aux critères normatifs comme les normes morales, les intentions et les motivations des agents humains. Contrairement aux modèles rationalistes, les auteurs du CCM considèrent que l'objectif des individus n'est pas seulement de comprendre le monde tel qu'il est, ils doivent également détecter les intentions malveillantes et mettre en place un système de blâme et de punitions. Ce second objectif induirait une motivation à blâmer certains individus pouvant biaiser l'application des critères normatifs entre leurs actions et le dommage. Néanmoins, cette hypothèse n'a jamais été réellement étudiée. Nous détaillerons comment ce modèle apporte une explication recouvrant à la fois l'effet de l'infraction d'une norme, mais également l'effet des préjugés ou encore des évaluations et émotions négatives, tout ceci qualifiant un effet plus large, à savoir celui de la valence morale de l'agent.

A partir de la base théorique présentée dans les deux premiers chapitres, deux axes de recherches ont été développés pour étudier la rationalité des jugements causaux. Le premier axe de recherche (Chapitre 3) visait à investiguer l'articulation entre l'effet de la valence morale de l'agent sur le jugement causal et les critères normatifs présentés dans le Chapitre 1. Nous avons notamment étudié (1) si les individus allaient jusqu'à inventer un lien causal inexistant entre l'agent et le dommage, et (2) si l'effet de la valence morale était modéré par le degré de force causale entre l'agent et le dommage.

Le deuxième axe de recherche (Chapitre 4) visait à aller encore plus loin dans le questionnement de la rationalité des jugements causaux de sens commun en adoptant une perspective pragmatique de l'effet de la valence morale. Cette dernière considère que les auteurs du CCM concluent à la présence d'un biais de jugement car leurs critères normatifs sont incomplets, à savoir qu'ils n'intègrent pas les critères normatifs liés aux règles de la conversation. En nous appuyant sur certaines théories et résultats pragmatiques, nous avons étendu cette analyse en mettant en évidence un problème méthodologique majeur des études ayant étudié l'effet de la valence morale. Nous avons également mené deux études expérimentales permettant de confronter les hypothèses du CCM avec celles de l'approche pragmatique du jugement causal.

Enfin, dans le Chapitre 5, nous synthétiserons les apports du présent travail de thèse, tant sur le plan théorique qu'appliqué. Nous discuterons de certaines limites théoriques et méthodologiques pouvant porter préjudice à nos conclusions et nous proposerons des perspectives de recherches futures.

Paradigme expérimental et précautions scientifiques

Pour répondre à nos objectifs de recherche, nous emploierons le paradigme expérimental en deux étapes le plus utilisé dans la littérature sur le jugement causal. Dans ce paradigme, les participant.e.s sont invité.e.s à lire un scénario sous forme de texte, présentant un ensemble d'événements impliquant un agent humain et aboutissant à un dommage. Après avoir lu le scénario, plusieurs questions sont posées vis-à-vis de l'événement, notamment une ou plusieurs questions de jugement causal. L'avantage principal de l'utilisation de questionnaires en ligne est d'obtenir un nombre important de réponses (1488 participants sur l'ensemble de nos études), permettant notamment d'obtenir une puissance statistique convenable.

Il nous importe de préciser que ce travail de thèse a pris en considération certains principes établis par la communauté scientifique suite à la crise de la réplication en psychologie, et notamment en psychologie sociale (25% de réplifications sur un total de 55 études en psychologie sociale, Open Science Collaboration, 2015). Conformément aux recommandations de Simmons et ses collaborateurs (2011), nous avons justifié la taille des échantillons utilisés dans nos études à partir d'analyses de puissance statistique a priori. Lorsque ces analyses n'étaient pas faisables dans certaines conditions en raison d'une absence de résultats préalables permettant d'anticiper une taille d'effet donnée (i.e., études 5 et 6), nous avons mené des analyses de sensibilité pour déterminer quelle taille d'effet minimale nous pouvions détecter de manière fiable avec notre échantillon (Lakens, 2022). Nous avons également souhaité nous inscrire dans une démarche de science ouverte en partageant les données anonymes sur la plateforme Open Science Framework (<https://osf.io>), et conformément au règlement général sur la protection des données (RGPD). Tous les fichiers de données anonymisées des études présentées dans le présent manuscrit sont téléchargeables à partir de ce lien : <https://osf.io/6yuhj/>.

Enfin, un autre aspect que nous avons pris en compte est l'incapacité des statistiques fréquentistes à réellement faire la distinction entre l'*absence de preuve* (i.e., les données recueillies ne permettent pas de conclure sur l'existence de l'effet) et la *preuve de l'absence* (i.e., l'effet n'existe pas au niveau de la population) (Wagenmakers et al., 2018). Les statistiques bayésiennes, quant à elles, permettent de quantifier le soutien envers l'hypothèse nulle et d'indiquer lorsque les données ne sont pas diagnostiques pour la départager de l'hypothèse alternative (Dienes, 2014 ; Van den Bergh et al., 2020). Cette thèse visait essentiellement à étudier les conditions d'apparition d'un effet, il nous était primordial de déterminer si nos résultats soutenaient l'absence de l'effet ou si nos données ne permettaient pas de conclure. Pour pallier cela, des facteurs bayésiens ont été utilisés pour tous les tests

concernant l'effet de la valence morale. Toutefois, ces facteurs bayésiens ont eu pour vocation de compléter les statistiques fréquentistes classiques, et non de les supplanter (Kelter, 2020 ; Quatto et al., 2020).

Évidemment, toutes ces mesures sont perfectibles. Nous regrettons rétrospectivement de ne pas avoir préenregistré nos hypothèses ainsi que nos protocoles. Nous tâcherons encore d'améliorer notre approche scientifique après ce travail de thèse.

Chapitre 1. Les critères normatifs du jugement causal

« The assertion that a cognitive bias is present is not so easy to make when there is no normative model specifying the expected response in the reference situation. »

Caverni, Fabre et Gonzales (1990, p.8)

Les modèles normatifs majoritairement utilisés dans la recherche en psychologie sont des modèles logiques ou mathématiques. Nous pouvons notamment citer la théorie des probabilités pour le biais de récence, le théorème de Bayes pour la négligence du taux de base ou encore le falsificationnisme poppérien pour le biais de confirmation (pour d'autres exemples de modèles normatifs utilisés en psychologie, voir Evans, 2007). Les travaux pionniers de Kahneman et Tversky (1979) portaient notamment sur des biais pour lesquels les mathématiques constituaient le modèle normatif. Or, si l'efficacité et la fiabilité des mathématiques n'est pas débattue, ce n'est pas le cas des modèles normatifs du jugement causal.

Déterminer un modèle normatif du jugement causal n'est pas aussi aisé que pour la plupart des tâches de raisonnement. En effet, il n'existe pas encore de consensus entre les spécialistes du domaine pour définir ce qu'est la causalité, et comment nous devrions établir de manière rationnelle des relations causales dans notre environnement (pour une revue des différentes approches, voir Beebee et al., 2009 ; White, 1990). Néanmoins, nous verrons que deux critères fondamentaux du jugement causal tant sur le plan philosophique et logique que sur le plan juridique peuvent être distingués. Ces deux critères seront détaillés dans ce chapitre. Ils constitueront le point de référence à partir duquel nous pourrions déterminer si le jugement causal de sens commun pour les événements à l'issue négative est biaisé. Nous commencerons

par définir brièvement ces critères d'un point de vue philosophique et logique. Nous verrons ensuite comment le droit s'est emparé de ces notions.

D'un point de vue philosophique et logique

Les deux approches philosophiques de la causalité

Les premières tentatives de définition de la causalité sont très anciennes. Aristote au IV^{ème} siècle avant l'ère commune, ou encore William d'Ockahm au 14^{ème} siècle avaient déjà proposé des définitions du mot *cause* (pour un détail et un historique de ces définitions, voir White, 1990). Néanmoins, ces définitions ne sont plus considérées par les philosophes contemporains et n'ont plus grand chose à voir avec le mot *cause* tel que nous l'employons dans le langage courant. Actuellement, les théories philosophiques de la causalité peuvent être sommairement divisées en deux catégories (Gerstenberg et al., 2021 ; Hall, 2004 ; Lagnado & Gerstenberg, 2017 ; Paul et al., 2013) : les *théories du processus* et les *théories de la dépendance*.

Selon les *théories du processus* (Dowe, 2000 ; Machamer et al., 2000 ; Wolff, 2007), deux objets A et B sont reliés causalement s'il existe un processus physique continu entre eux, tel qu'un transfert de force ou une réaction chimique. Par exemple, nous pouvons déterminer que le poison versé par Monsieur X dans le verre de Monsieur Y est la cause de la mort de ce dernier, car il y a eu un processus physique et chimique complexe entre l'ingestion de poison et la mort de Monsieur Y. Bien que plutôt intuitives, ces théories présentent un certain nombre de limites. La principale limite est celle des cas d'omission ou d'absence d'action. Imaginons des parents maltraitants qui ne nourrissent plus leur enfant, et que l'enfant décède. Il est difficile d'imaginer que ces parents ne puissent pas être considérés comme une cause de la mort de l'enfant. Néanmoins, il n'y a aucun processus physique entre les deux éléments, c'est justement l'absence de processus physique (i.e., donner de la nourriture à son enfant) qui qualifie ici la

CHAPITRE 1

LES CRITÈRES NORMATIFS DU JUGEMENT CAUSAL

relation causale. Dans la vie quotidienne, nombreux sont les exemples dans lesquels une omission est considérée comme une cause : un pilote d'avion qui ne sort pas les trains d'atterrissage, un plongeur qui oublie de remplir sa bouteille d'oxygène, un mari qui ne dit pas à sa femme que les freins de sa voiture son défectueux, etc.

Selon les *théories de la dépendance* (Halpern, 2016 ; Hitchcock, 2007 ; Hume, 1748 ; Lewis, 1973, 2000 ; Mackie, 1980 ; Woodward, 2006), A et B sont reliés causalement si l'existence de B dépend de l'existence de A. Autrement dit, A doit être une condition nécessaire de B pour être qualifié de cause. Par exemple, si Monsieur Y n'était pas mort sans l'action de Monsieur X, alors Monsieur X est la cause de la mort de Monsieur Y. Cette approche est également nommée la conception *contrefactuelle* de la causalité. Dans l'exemple des parents maltraitants, leur action est bien une cause de la mort de l'enfant puisque si les parents n'avaient pas cessé de nourrir leur enfant, alors celui-ci ne serait pas mort. Leur omission est donc une condition nécessaire au décès de l'enfant. De plus, la conception contrefactuelle permet de distinguer la corrélation de la causalité de la manière suivante : si A cause B, un changement dans A produit un changement dans B mais l'inverse n'est pas vrai (contrairement à une corrélation). Enfin, outre le fait de répondre au problème des cas d'omission et de la simple corrélation, un autre avantage des théories de la dépendance est qu'elles permettent une définition à ce point précise et efficace de la causalité que certains auteurs l'ont traduite en formalisme logique (Halpern, 2016 ; Halpern & Pearl, 2005). Il s'agit de la conception de la causalité la plus partagée. C'est à partir de cet ancrage philosophique que la plupart des modèles normatifs du jugement causal appliqués à l'intelligence artificielle (e.g., Andreas et al., 2022 ; Pearl, 2000), à la psychologie (e.g., Halpern & Pearl, 2005 ; Waldmann & Hagmayer, 2013) et aux systèmes juridiques français et anglo-saxons (e.g., Stapleton, 2008, 2015) se sont construits. Pour ces raisons, le présent travail de thèse s'inscrit également dans cette définition

de la causalité. Nous allons désormais aborder les deux critères normatifs particulièrement importants pour cette conception contrefactuelle : la nécessité réelle et la sensibilité.

De la nécessité à la nécessité réelle.

Comme nous l'avons vu précédemment, une cause peut être définie comme une condition nécessaire à un événement. Cependant, une précision importante a été apportée par la suite. De nombreux auteurs ont en effet relevé le problème de la *surdétermination* et de la *préemption* (Halpern & Pearl, 2005 ; Hitchcock, 2007 ; Lewis, 1973, 2000 ; Pearl, 2000). La *surdétermination* a lieu lorsque plusieurs conditions sont chacune suffisante pour provoquer le dommage. La *préemption*, quant à elle, est une situation dans laquelle une condition aurait provoqué le dommage si une autre condition ne l'avait pas causé avant. Reprenons l'exemple de Mandel (2003) : un criminel professionnel est la proie de deux tentatives d'assassinat, une par l'assassin 1 et une autre par l'assassin 2. L'assassin 1 donne au criminel professionnel une dose létale de poison qui la tuera en une heure. Avant que le poison ne fasse effet, l'assassin 2 percute la voiture de la victime et la fait sortir de la route. La voiture explose et la victime est tuée. Dans cet exemple, soutenir que l'assassin 2 a causé la mort de la victime paraît plutôt évident. Pourtant, son action n'est pas une condition nécessaire au décès de la victime puisque celle-ci serait advenue quoi qu'il en soit à cause du poison (i.e., lui-même suffisant pour tuer la victime).

Il existe plusieurs solutions à ce problème (Gerstenberg et al., 2021). La principale solution est d'augmenter le niveau de granularité du résultat (Lewis, 2000 ; Mandel, 2003 ; Schaffer, 2005). Dans cet exemple, il s'agit de considérer que l'assassin 2 est une condition nécessaire à la manière exacte dont la victime est décédée. L'idée est donc de considérer le résultat tel qu'il est survenu⁴ (e.g., l'explosion de la voiture telle qu'elle a réellement eu lieu),

⁴ « the effect as it came about » (Mackie, 1980, p.40)

plutôt qu'un résultat dans l'absolu (e.g., la mort de la victime dans n'importe quelles circonstances). C'est ce que les auteurs ont nommé la *causalité réelle*⁵ (Halpern, 2016). La causalité réelle (e.g., la cigarette a provoqué le cancer de Paul) doit ainsi être distinguée de la causalité générale (e.g., la cigarette cause le cancer) (Lagnado et al., 2013). Cette nuance est fondamentale car, contrairement à la recherche scientifique qui vise à établir des lois universelles, lorsque le sens commun ou le monde judiciaire cherchent à expliquer un dommage, c'est bien le dommage tel qu'il est survenu qui intéresse (Hart & Honoré, 1985 ; Mackie, 1980 ; Woodward, 2003). En effet, lorsqu'une personne décède, nous ne souhaitons pas connaître tous les facteurs qui auraient pu la tuer, mais plutôt le facteur qui a effectivement provoqué son décès dans ces circonstances précises. Par conséquent, le critère normatif que nous devons prendre en compte pour évaluer la rationalité des jugements causaux de sens commun (ou judiciaires) n'est pas la seule nécessité mais la *nécessité réelle* (e.g., Icard et al., 2017). Ainsi, le jugement causal d'un individu pourra être considéré comme rationnel si celui-ci sélectionne une condition nécessaire réelle de l'événement à expliquer (Halpern & Hitchcock, 2015 ; Halpern & Pearl, 2005). Si la cause qu'il mentionne n'est pas une condition nécessaire réelle au dommage à juger, alors celui-ci viole la règle fondamentale de la causalité et son jugement peut être qualifié de *biaisé*.

D'une causalité dichotomique à une causalité graduée

La notion de *nécessité réelle* implique une conception dichotomique de la causalité : un élément est une condition nécessaire réelle ou n'est pas une condition nécessaire réelle. Bien que plusieurs philosophes considèrent uniquement ce critère pour définir la causalité (e.g., Hall, 2004 ; Lewis, 1973 ; Mill, 1843), cette définition n'est pas suffisamment discriminante pour le sens commun. La raison est la suivante : chaque événement possède un nombre de conditions

⁵ *actual causation*

CHAPITRE 1

LES CRITÈRES NORMATIFS DU JUGEMENT CAUSAL

nécessaires réelles qui dépasse vraisemblablement les capacités de traitement des individus. En effet, chaque action met en mouvement des molécules qui vont affecter d'autres molécules, qui affectent à leur tour d'autres molécules, et ainsi de suite (chaque mouvement de molécule étant une condition nécessaire réelle aux mouvements des suivantes)⁶. C'est pourquoi, afin d'affiner cette définition trop inclusive de la causalité, des philosophes ont proposé une conception graduée de la causalité (e.g., Demirtas, 2022 ; Halpern & Hitchcock, 2015 ; Usher, 2020).

Le philosophe James Woodward (2006) a notamment introduit le concept de *sensibilité*, parfois nommé *suffisance robuste* (e.g., Grinfeld et al., 2020 ; Icard et al., 2017) ou *stabilité* (Vasilyeva et al., 2018), pour qualifier la force d'une relation causale. Dans cette perspective, une relation causale est faible lorsque celle-ci est sensible aux circonstances de fond. Halpern (2016) illustre ce concept avec l'exemple suivant : une petite fille sera davantage causalement reliée à la destruction d'une bouteille si elle jette une pierre dessus et la brise plutôt que si elle « jette une pierre sur une serrure, qui s'ouvre, ce qui fait que le lion qui était dans la cage fermée s'échappe, effrayant le chat, qui bondit sur la table et renverse la bouteille, qui se brise alors. » (p.53). Bien que la petite fille soit une condition nécessaire réelle de la destruction de la bouteille dans les deux cas, le lien causal du deuxième scénario dépend de nombreuses circonstances (e.g., la solidité de la serrure de la cage, le comportement du lion, la présence du chat, etc.). En d'autres termes, le lancer de la petite fille est beaucoup plus sensible aux circonstances de fond dans le second cas, ce qui implique un lien causal plus faible que le lancer décrit dans le premier cas. La sensibilité permet ainsi de ne pas traiter toutes les conditions nécessaires réelles de la même manière en attribuant différents degrés d'importance (ou force causale) à chacune d'entre elles.

⁶ L'expression *effet papillon* permet d'illustrer cela : théoriquement, les battements d'ailes d'un papillon au Brésil pourrait provoquer une tornade au Texas par le biais d'une chaîne causale complexe (Lorenz, 1972).

Les deux critères de présence et de force d'un lien causal permettent conjointement d'établir le *modèle causal* (ou *réseau bayésien*) d'une situation que nous allons voir à présent.

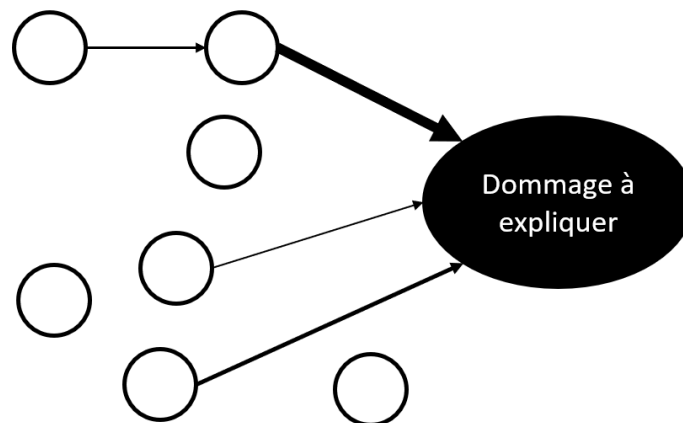
Le modèle causal d'une situation

La notion de modèle causal est apparue à la fin des années 80 en intelligence artificielle⁷ dans le but de créer des systèmes informatiques experts pouvant prendre des décisions dans des situations d'incertitude (Drouet, 2007). Il s'agissait donc d'une approche normative dont l'objectif était de concevoir le système le plus rationnel et efficace possible. En cela, les modèles causaux peuvent être considérés comme le modèle normatif du jugement causal. Concrètement, un modèle causal est un « moyen visuel compact de représenter les relations causales entre les variables » (Rottman, 2017, p.165). Il est ainsi généralement représenté sous forme graphique comme l'illustre la Figure 1 (p.20). Dans ces graphes, les cercles (ou *nœuds*) représentent les variables et les flèches représentent la présence d'une relation causale entre ces variables (Cooper, 1999 ; Glymour & Cheng, 1998 ; Griffiths & Tenenbaum, 2005). L'absence d'une flèche indique l'absence de relation causale entre les variables. La taille des flèches, quant à elle, représente la force causale de chaque relation : plus une flèche est épaisse, plus la relation causale est forte. Ceci permet notamment d'identifier la cause qui a le plus contribué au résultat.

⁷ A noter que les travaux en intelligence artificielle se basent également sur une définition contrefactuelle de la causalité (Andreas et al., 2022). C'est pourquoi on parle de *graph surgery* pour désigner les interventions contrefactuelles menées sur les graphes (Pearl, 2000 ; Waldmann & Hagmayer, 2013).

Figure 1

Exemple de modèle causal d'une situation à l'issue négative.



En ce qui concerne la notion de force causale, différentes conceptions existent (Perales et al., 2017). Les réseaux bayésiens la quantifient généralement par une probabilité : chaque cause va augmenter plus ou moins la probabilité que l'effet survienne (e.g., Hagmayer, 2016 ; Lu et al., 2008 ; Pearl, 1999). Or, bien que proches et permettant de rendre compte de la force d'une relation causale, les concepts de sensibilité et de probabilité ne sont pas identiques et la notion de probabilité présente un certain nombre de limites (pour plus de détails, voir Grinfeld et al., 2020). Le principal avantage du concept de sensibilité sur celui de probabilité est que, généralement, les individus ont accès aux circonstances de fond qui modèrent la relation causale (i.e., les autres éléments ayant contribué au résultat). En revanche, ils ont rarement accès à des informations fiables pour porter des jugements sur l'augmentation ou la diminution de la probabilité du résultat en fonction de telle ou telle condition. Bien que l'utilisation de probabilités pour quantifier la force d'une relation causale ne puisse pas être considérée comme un critère normatif totalement inapproprié, nous nous centrerons sur le critère de sensibilité dans ce travail de thèse pour les raisons évoquées précédemment. De plus, la relative simplicité du critère de sensibilité par rapport à celui de probabilité le rend davantage implémentable à la psychologie humaine.

En résumé, bien que le débat sur une définition philosophique de la causalité ne soit pas tranché, la conception contrefactuelle (i.e., *théories de la dépendance*) semble présenter moins de limites que les *théories du processus* puisque les situations de préemptions et de surdétermination peuvent être surmontées grâce au principe de causalité réelle. De plus, cette conception contrefactuelle permet de dégager deux critères normatifs précis : la *nécessité réelle* (caractérisant la présence ou l'absence d'un lien causal) et la *sensibilité* (caractérisant la force de ce lien). Ces deux critères permettent de construire le modèle causal d'une situation, c'est-à-dire une représentation complète⁸ et rationnelle des liens causaux d'une situation donnée. Toutefois, l'approche théorique des modèles causaux est un domaine de recherche actif plutôt qu'une théorie fixe et unifiée (Rottman, 2017). A l'heure actuelle, il n'existe pas de cadre normatif définitif du jugement causal. Néanmoins, bien qu'améliorable et en cours de développement, le cadre normatif présenté dans cette partie peut être qualifié de rationnel (Holyoak & Cheng, 2011). Ainsi, lorsque le jugement d'un individu ne respecte pas le modèle causal d'une situation décrit dans cette partie, alors celui-ci pourra être qualifié de *biaisé* ou *d'irrationnel* (conformément à la notion de rationalité normative dans laquelle s'inscrit cette thèse). Cela peut se produire si l'individu désigne une condition réelle non nécessaire comme cause ou s'il désigne, comme cause principale, une cause qui ne possède pas la plus grande force causale (i.e., une condition qui n'est pas la moins sensible aux circonstances).

Le cadre normatif de cette thèse a été défini en déterminant comment l'individu devrait penser pour produire un jugement causal rationnel. Néanmoins, l'approche philosophique décrit une rationalité dans l'absolu, c'est-à-dire comment un individu devrait penser la causalité dans n'importe quelles circonstances. Or, dans les situations à l'issue négative impliquant un

⁸ Le modèle causal est considéré comme complet conformément au principe de Markov selon lequel les variables parentes à celles du modèle (i.e., les variables ayant causé celles du modèle) n'apportent pas davantage d'information que celles qui sont déjà dans le modèle (voir Pearl, 2000 ; Sloman et al., 2009).

agent humain, le jugement causal de sens commun est étroitement lié aux jugements de responsabilité et de blâme (Hart & Honoré, 1985). Par ailleurs, le principal domaine d'application de cette thèse étant dans le domaine judiciaire, il est essentiel de définir brièvement le jugement causal normatif spécifique à ce domaine.

D'un point de vue juridique

Comme nous l'avons évoqué en introduction, le jugement causal est l'un des processus psychologiques les plus fondamentaux pour les verdicts judiciaires. La première étape avant de juger et de sanctionner un individu est d'établir un lien causal entre ses actions et le dommage qui lui est reproché (Hart & Honoré, 1985). Comme pour l'approche philosophique, déterminer un cadre normatif juridique pour le jugement causal n'est pas aisé. La question se pose de savoir si les critères normatifs prescrits par l'approche philosophique peuvent être directement appliqués au domaine judiciaire. Selon Lagnado et Gerstenberg (2017), deux approches sont à distinguer : certains auteurs soutiennent que le droit fonctionne avec sa propre notion de causalité et que les conceptions philosophiques sont souvent non pertinentes (Green, 2015 ; Hoffman, 2011). D'autres auteurs, en revanche, plaident pour une théorie générale de la causalité qui pourrait s'appliquer dans différents domaines tels que le droit, la science et la métaphysique (Moore, 2009 ; Schaffer, 2010). Bien que ce débat reste ouvert à ce jour, nous verrons que les deux critères philosophiques vus précédemment trouvent leur parallèle dans les théories juridiques de la causalité.

La présomption de rationalité

Soulignons tout d'abord la nécessité de construire et de poser un cadre normatif dans ce domaine particulier que constitue la justice. En effet, de nombreux systèmes juridiques se sont construits sur la notion de jury populaire. Ces jurys peuvent être composés de profanes et de professionnels du droit (e.g., comme en France, Italie, Allemagne, Belgique, etc.) ou

essentiellement de profanes (e.g., comme en Angleterre, Pays de Galle, Etats-Unis, Australie, Canada, etc.) pour juger des affaires criminelles graves (Kapardis, 2009). Ce modèle juridique réclame que la justice soit rendue au nom du peuple et que les criminels soient jugés par leurs pairs (Gissinger-Bosse, 2012). Le jury populaire est ainsi le garant des « libertés individuelles et tend à donner confiance dans le fonctionnement de la justice » (Berger, 2008, p.24).

Si le rôle actif du citoyen dans l'institution judiciaire permet de revendiquer et de maintenir une certaine idéologie politique (i.e., la souveraineté du peuple), cette idéologie semble toutefois reposer sur une conception « naïve » du raisonnement humain, en partant d'une présomption de rationalité de la part des juges. En effet, ce système a été pensé et mis en place à la fin du 18^{ème} siècle, soit bien avant les premiers travaux en psychologie sur les erreurs de raisonnement de sens commun (e.g., Kahneman & Tversky, 1979 ; Wason, 1960). Cette notion de biais, c'est-à-dire d'écart entre le cadre normatif approprié et les jugements rendus par les jurys populaires, n'a donc pas été prise en compte lors de la construction de nos systèmes judiciaires. Or, de nombreux travaux récents ont démontré l'impact de facteurs extra-normatifs ou extra-légaux (pour une revue, voir Devine et al., 2001, voir aussi Nadelhoffer, 2006), tels que l'origine ethnique (Anwar et al., 2012 ; Mitchell, 2005) ou l'attractivité physique de l'accusé (e.g., Efran, 1974 ; Saladin et al., 1988) sur les verdicts judiciaires. Bien que le jury populaire soit moins employé en France pour des raisons de contraintes temporelles et de fluidité des procédures⁹, les jugements des professionnels du droit sont loin d'être totalement impartiaux et objectifs (pour une revue, voir Peer & Gamliel, 2013).

Étant donné l'importance des enjeux, il est primordial de déterminer dans quelle mesure les jugements de sens commun (comme celui des magistrats professionnels) s'écartent du

⁹ Depuis le projet de loi du 22 décembre 2021 pour la confiance dans l'institution judiciaire (modifiant l'article 181 du code de procédure pénale) : les crimes punissables d'une peine pouvant atteindre 20 ans de réclusion seront jugés par une « cour criminelle » composée uniquement de professionnels du droit.

jugement causal du « juré idéal ». A nouveau, pour déterminer cet écart, la première étape est de décrire ce que serait ce jugement causal idéal dans ce domaine précis.

Déterminer un lien causal entre les actions de l'accusé et le dommage

Pour déterminer la responsabilité d'un accusé, le juré doit établir la dimension objective et la dimension subjective de ses actes (Bordel, 2002 ; Bordel et al., 2006 ; Fauconnet, 1920). La responsabilité objective renvoie au lien causal entre les actions de l'accusé et le dommage (e.g., l'accusé a-t-il commis un acte qui a provoqué le décès de la victime ?). La responsabilité subjective, quant à elle, renvoie aux états mentaux de l'accusé (e.g., ses intentions et la justification de ses actions au moment des faits). Cette distinction entre lien causal et états mentaux de l'accusé peut être retrouvé dans le système juridique anglo-saxon à travers les notions d'*actus reus* et de *mens rea* (Finkel & Groscup, 1997) et celles d'*élément matériel* et d'*élément moral* dans le système juridique français (Bordel et al., 2006). Nous ne définirons pas le cadre normatif du jugement de responsabilité subjective ici. C'est la responsabilité objective qui va nous intéresser dans le cadre du jugement causal à travers la question suivante : Comment établir un lien causal entre les actions de l'accusé et le dommage conformément aux attentes du droit ? Formulée autrement, quel est le cadre normatif du jugement causal pour le droit ?

Bien que le jugement causal soit central pour établir cette dimension objective, le cadre normatif de celui-ci n'est pas aussi bien défini que dans le domaine philosophique. La définition du cadre normatif pour le jugement causal judiciaire a par ailleurs suscité de nombreux débats (Güver & Kneer, 2022 ; Lagnado & Gerstenberg, 2017). Néanmoins, la philosophie et le droit peuvent tous deux s'accorder sur une conception contrefactuelle de la causalité. En effet, le système juridique français¹⁰ est, comme l'approche philosophique,

¹⁰ Tout comme les systèmes juridiques anglo-saxons (voir Andreas et al., 2022 ; Stapleton, 2015)

CHAPITRE 1

LES CRITÈRES NORMATIFS DU JUGEMENT CAUSAL

porteur d'une définition contrefactuelle de la causalité plutôt que d'une définition en termes de processus. Ainsi, les omissions et les négligences peuvent être sanctionnées et donc considérées comme causalement reliées au dommage. L'article 1241 du Code civil illustre explicitement ceci : « Chacun est responsable du dommage qu'il a causé non seulement par son fait, mais encore par sa négligence ou par son imprudence. ». Le code pénal mentionne également les absences d'action. Par exemple, l'article 223-6 du code pénal français sanctionne la non-assistance à personne en danger ou l'article 121-3 de ce même code qui sanctionne la négligence. A l'instar des théories de la dépendance, l'absence d'action (e.g., omission, négligence) peut constituer la cause d'un dommage à condition que cette absence soit une condition nécessaire réelle à la survenue de celui-ci. Nous constaterons que, en plus de la conception contrefactuelle, les deux critères normatifs précédemment décrits (i.e., nécessité réelle et sensibilité) peuvent trouver leurs équivalents dans les normes juridiques. Cette constatation nous offre ainsi la possibilité d'un cadre normatif unifié entre la philosophie et le droit.

Concernant les critères normatifs, deux théories juridiques sont couramment utilisées en France : la *théorie de l'équivalence des conditions* et la *théorie de la causalité adéquate* (Nasri, 2006 ; Quézel-Ambrunaz, 2008). Selon la théorie de l'équivalence des conditions, tous les faits sans lesquels le dommage ne se serait pas produit sont considérés comme en étant la cause. Ceci renvoie à la notion de *condition nécessaire réelle* vue précédemment : une cause (C) est une condition nécessaire réelle de l'événement (E) si, en l'absence de C, E ne se produit pas (Icard et al., 2017). Il s'avère que le droit anglo-saxon s'appuie également sur ce critère et cette conception contrefactuelle à travers le principe *sine qua non* : un accusé ne peut être déclaré coupable d'un dommage que si ses actions constituent une condition *sine qua non* de celui-ci (Hart & Honoré, 1985 ; Stapleton, 2008, 2015 ; Summers, 2018). La notion de cause réelle a également été intégrée dans la conception juridique de la causalité (Halpern, 2005 ;

CHAPITRE 1

LES CRITÈRES NORMATIFS DU JUGEMENT CAUSAL

Wright, 1985). En effet, un accusé ne sera déclaré coupable d'un dommage que si celui-ci l'a effectivement produit. Pour reprendre l'exemple de Mandel (2003), l'individu qui a empoisonné la victime décédée dans un accident de voiture avant que le poison ne fasse effet pourra être jugé coupable de tentative d'homicide mais pas d'homicide. C'est pourquoi le critère de nécessité réelle (et non pas de simple nécessité) est également la pierre angulaire du jugement causal normatif pour les prises de décisions judiciaires. Comme nous l'avons vu avec l'approche philosophique, la principale limite à cette théorie est liée à la complexité des événements et à la ramification des conséquences qui peuvent être associées même au plus simple des comportements. Étant donné que chacun de nos comportements produit théoriquement des chaînes causales infinies (Mischel, 1970 ; White, 1968), nous serions toutes et tous responsables d'un nombre incalculable de dommages pour lesquels nous sommes des conditions nécessaires réelles. Imaginons un conducteur en excès de vitesse qui heurte un individu et le blesse. L'individu est transporté à l'hôpital et reçoit du sang contaminé à l'hépatite C à cause d'un employé de laboratoire négligent. Dans cette situation, le conducteur sera très probablement déclaré coupable des blessures provoquées par l'accident. En revanche, il est également une condition nécessaire réelle à la contamination sanguine puisque, sans son action, le sang de l'individu n'aurait pas été contaminé. La question est : doit-on considérer le conducteur comme coupable de la contamination sanguine en plus des blessures infligées par l'accident ? Selon la théorie de l'équivalence des conditions, la réponse est oui. Selon la théorie de la causalité adéquate, la cause de la contamination se limite à la condition la plus déterminante pour le dommage (i.e., l'employé de laboratoire).

Si la théorie de la causalité adéquate a l'avantage de proposer une définition plus restrictive de la causalité, la notion de « condition la plus déterminante » reste en revanche peu précise. Fumerton et Kress (2001) évoquent notamment la notion de cause proximale : « Une cause immédiate est typiquement définie comme une cause qui produit un dommage dans une

CHAPITRE 1

LES CRITÈRES NORMATIFS DU JUGEMENT CAUSAL

séquence naturelle, directe et continue, sans aucune cause de substitution qui rompt la chaîne causale. » (p.87). Si cette notion de proximité reste peu précise, elle fait écho à la notion de *sensibilité* de l'approche philosophique permettant de définir la *force causale* d'une condition nécessaire réelle. Dans notre exemple, l'action du conducteur n'est pas une condition suffisamment robuste de la contamination du sang de la victime car si l'employé de laboratoire n'avait pas été négligent, cette contamination n'aurait pas eu lieu. En d'autres termes, l'action du conducteur est sensible aux circonstances de fond pour produire la contamination sanguine, et ce, contrairement à l'action de l'employé de laboratoire. Conformément à la théorie de la causalité adéquate, seul l'employé de laboratoire serait désigné coupable de cette contamination. En somme, le conducteur est une condition nécessaire réelle à ce dommage, mais il ne possède pas une force causale (insensibilité) assez importante pour être désigné coupable. Le droit anglo-saxon inclut également cette notion de force causale. Par exemple, dans l'Etat de New-York, si une personne provoque un incendie de manière non intentionnelle, elle est responsable de la première maison brûlée par l'incendie mais pas des autres maisons (Alicke, 1992). Il n'est pas étonnant de constater que les systèmes juridiques incluent la notion de force causale en plus de celle de nécessité réelle car, s'il en était autrement, le conducteur de notre exemple serait non seulement coupable de la contamination sanguine mais également de toutes les autres conséquences qui pourraient suivre : la dépression de la conjointe de la victime, la perte du chiffre d'affaire de l'entreprise dans laquelle travaillait la victime, l'échec scolaire des enfants de la victime, etc. L'application des peines et l'impossibilité de prévoir l'ensemble des conséquences de nos actions ont donc naturellement incité nos sociétés à mettre en place des critères vis-à-vis de la causalité : une action doit non seulement être une condition nécessaire réelle mais également suffisamment insensible aux circonstances environnantes.

Si l'application de ces deux critères est nécessaire pour déterminer un lien causal entre un accusé et un dommage, il est également important que cette application s'opère de façon

CHAPITRE 1

LES CRITÈRES NORMATIFS DU JUGEMENT CAUSAL

exclusive, c'est-à-dire qu'aucun autre critère ne doit intervenir dans l'évaluation de ce lien causal. C'est la distinction qu'opèrent Fumerton et Kress (2001) entre les facteurs causaux légalement pertinents et légalement non pertinents. Les intentions de l'accusé, sa personnalité ou encore son attractivité sociale ne doivent pas influencer cette évaluation. Les intentions et la personnalité peuvent/doivent intervenir dans le jugement du *mens rea* (ou *élément moral*) mais en aucun cas dans celui de l'*actus reus* (ou *élément matériel*). En d'autres termes, les réactions affectives et les évaluations morales envers l'accusé ne doivent pas intervenir dans cette analyse du lien causal. En somme, le droit attend une analyse séquentielle et rationnelle des faits. Premièrement, le jury doit déterminer un lien causal entre les actions de l'accusé et le dommage par l'utilisation exclusive des deux critères présentés. Deuxièmement, il doit déterminer l'élément moral permettant d'établir la responsabilité de l'accusé (e.g., L'accusé possède-t-il toutes ses facultés mentales ? Avait-il l'intention de commettre le dommage ? A-t-il été négligent ?). Cette analyse peut être résumée de la manière suivante : jugement causal → jugement moral → attribution de sanctions.

En résumé, les décisions judiciaires peuvent être biaisées de deux façons : (1) la non-prise en compte des critères normatifs du droit ; (2) l'utilisation d'autres critères que ceux imposés par la loi (von Helversen & Rieskamp, 2009). Nous avons vu que les critères normatifs du droit concernant le jugement causal peuvent être assimilés à ceux de l'approche philosophique : la *nécessité réelle* (présence d'un lien causal entre les actions de l'accusé et le dommage) et la *sensibilité* (ce lien causal doit présenter un certain niveau de force causale). Aucun autre critère ne doit être pris en compte dans le jugement causal. Le respect de ce cadre est nécessaire pour une application des peines juste et impartiale. En revanche, rien ne garantit que le jugement causal de sens commun le respecte scrupuleusement. C'est ce que nous étudierons dans le prochain chapitre.

Conclusion

Il s'agissait ici de définir le cadre normatif du jugement causal. Nous avons vu que ce cadre s'inscrit dans une conception contrefactuelle de la causalité et permet de distinguer deux critères, à savoir la nécessité réelle et la sensibilité. Le premier critère a été identifié très tôt par les philosophes et logiciens, et renvoie à la présence ou l'absence d'un lien causal entre deux événements. La non prise en compte de ce critère constitue l'infraction la plus « grave » du jugement causal. En effet, désigner comme cause une condition réelle non nécessaire reviendrait à inventer un lien causal inexistant. Dans un cadre judiciaire, cela consiste à déclarer coupable un accusé innocent, c'est-à-dire un accusé n'ayant pas causé le dommage qui lui est reproché. Si ce critère est fondamental, nous avons vu que toutes les conditions nécessaires réelles ne se valent pas (e.g., Les grands-parents d'un meurtrier ne sont pas responsables du meurtre de leur descendant). Ceci renvoie au second critère ayant été identifié plus tardivement. Ce critère est celui de la force d'un lien causal lorsque celui-ci est présent. Ce critère permet d'envisager la causalité, non pas comme dichotomique, mais comme graduelle. Le respect de ce critère permet un jugement causal cohérent et pratique en sélectionnant les facteurs les plus pertinents. Ces deux critères sont qualifiés de normatifs car c'est leur utilisation exclusive qui permet de qualifier un raisonnement de rationnel, tant d'un point de vue philosophique que judiciaire. Tout au long de cette thèse, ces deux critères serviront ainsi de point de référence permettant d'évaluer la rationalité des jugements causaux de sens commun. D'autre part, si nous observons des déviations systématiques dans les jugements par rapport à ces critères, ces derniers pourraient définir l'objectif à atteindre pour les travaux sur le débiaisement. Nous pouvons ainsi dégager une première grande question : Le jugement causal de sens commun est-il conforme aux critères présentés dans ce premier chapitre ?

Chapitre 2. De l'effet des normes à la validation du blâme

« The dream of constructing a theory that is acceptable both descriptively and normatively appears unrealizable. »

Tversky et Kahneman (1986, p.88)

Alors que la philosophie et les normes juridiques ont pour but de définir ce qu'est la causalité et comment nous devrions la penser, la psychologie s'attache à décrire et à expliquer comment les individus la pensent effectivement. Bien que ces deux approches ne remplissent pas la même fonction, les premiers modèles en psychologie ont été considérablement influencés par les modèles normatifs en philosophie (White, 1990 ; voir aussi Waldmann & Hagmayer, 2013). Cette tendance initiale de la recherche s'est traduite par une conception idéalisée de la rationalité des jugements de sens commun. En effet, les premiers modèles d'attribution causale évoqués en Introduction générale (e.g., Heider, 1958 ; Kelley, 1973) étaient construits à partir des règles de logique inductive de John Stuart Mill (1843) et de l'analyse de variance. L'individu est considéré comme un être rationnel qui, comme le scientifique, va chercher à « saisir la réalité » dans le but de la prédire et de la contrôler (Heider, 1958, p.79). L'analogie entre le sens commun et le scientifique est telle que ce courant de recherche a été baptisé le courant de « l'Homme comme scientifique » (Fincham & Jaspars, 1980). Fritz Heider, considéré comme le fondateur de ce courant, ira jusqu'à écrire que « la psychologie scientifique a beaucoup à apprendre de la psychologie de sens commun » (1958, p.5). De nombreuses critiques ont émergé par la suite et ont remis en question la capacité de ces modèles à prédire la manière dont les individus pensent réellement la causalité (pour une revue, voir Alicke et al., 2015). La principale critique adressée aux modèles du courant « l'Homme comme scientifique » est qu'ils seraient davantage prescriptifs que descriptifs : ils

décrieraient comment les individus devraient penser plutôt que comment ceux-ci pensent réellement (Alicke, 2011). Deux camps semblent toujours s'opposer sur ce point : les modèles défendant un individu rationnel (e.g., Kominsky et al., 2015 ; Gerstenberg et al., 2021) et ceux prétendant que le jugement causal des individus s'écarte systématiquement des critères normatifs (e.g., Alicke, 2000). Au cœur de ce débat, se trouve actuellement un effet largement répliqué : l'effet de l'infraction d'une norme morale sur le jugement causal. Nous allons voir dans ce chapitre que cet effet a été interprété comme l'application pertinente d'un critère de jugement causal par certains auteurs, et comme la manifestation d'un jugement biaisé par d'autres.

Effet de l'infraction d'une norme morale

Dans le premier chapitre, nous avons vu que, pour être rationnel, l'individu devait exclusivement utiliser deux critères. Or, de nombreux travaux en psychologie ont observé l'influence d'un facteur extra-normatif sur le jugement causal (e.g., Hitchcock & Knobe, 2009 ; Icard et al., 2017 ; Kominsky et al., 2015 ; Kominsky & Phillips, 2019 ; Samland & Waldmann, 2016). Ce facteur est l'infraction d'une norme morale (ou prescrite) : lorsqu'un agent enfreint une norme morale, celui-ci est davantage désigné comme cause du dommage par rapport à un agent n'ayant pas enfreint cette norme. Pour illustrer cet effet, reprenons le célèbre exemple de la *pen vignette* de Knobe et Fraser (2008), que nous utiliserons comme fil rouge dans ce manuscrit. Les auteurs ont fait lire le texte suivant aux participants :

« La réceptionniste du département de philosophie stocke les stylos dans son bureau. Les assistants administratifs sont autorisés à prendre des stylos, mais les membres du corps enseignant sont censés acheter les leurs.

Les assistants administratifs prennent régulièrement des stylos. Malheureusement, c'est aussi le cas des membres du corps enseignant. La réceptionniste leur a rappelé à plusieurs reprises par e-mail que seuls les assistants administratifs sont autorisés à en prendre.

Le lundi matin, l'un des assistants administratifs rencontre le professeur Smith qui passe devant le bureau de la réceptionniste. Tous deux prennent un stylo. Plus tard dans la journée, la réceptionniste doit prendre un message important... mais elle a un problème. Il n'y a plus de stylos sur son bureau. »

Les participants devaient ensuite indiquer s'ils étaient d'accord ou non avec chacune des phrases suivantes :

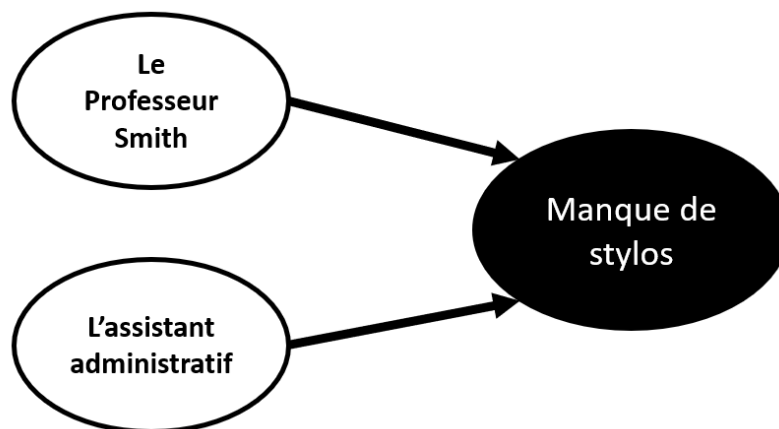
« Le professeur Smith a causé le problème. »

« L'assistant administratif a causé le problème. »

Knobe et Fraser (2008) ont constaté que les participants considéraient majoritairement le professeur Smith comme la cause du manque de stylos. Or, si nous reprenons les deux critères normatifs précédemment décrits, cette préférence en termes de sélection causale pour le professeur Smith ne semble pas rationnelle. Premièrement, les deux individus sont des conditions nécessaires réelles puisque sans l'action de l'un d'eux, la réceptionniste aurait eu un stylo. De plus, leur force causale est équivalente puisque leur action est identique : ils prennent chacun un stylo. Par conséquent, le modèle causal normatif de ce scénario est une *structure causale conjonctive* (Icard et al., 2017 ; Kominsky et al., 2015), c'est-à-dire un modèle causal non discriminant pouvant être illustrée par la Figure 2 ci-dessous.

Figure 2

Graphe du modèle causal de la pen vignette.



La seule différence entre les deux agents réside dans le fait que le Professeur Smith a enfreint une norme prescrite : le corps enseignant n'a pas le droit de prendre des stylos contrairement au personnel administratif. La prise en compte de l'infraction d'une norme prescrite (ou morale) dans le cadre du jugement moral ou de l'attribution de sanctions est parfaitement rationnel (e.g., Cushman, 2008 ; Malle et al., 2014 ; Rai & Fiske, 2011). En revanche, cet effet ne l'est pas dans le cadre du jugement causal. Pourtant, il a été répliqué un grand nombre de fois. Par conséquent, la question aujourd'hui n'est plus de savoir si cet effet existe ou non, mais d'en connaître sa nature et ses conditions d'apparition. C'est sur ce point que les modèles explicatifs actuels du jugement causal vont s'opposer. Nous allons désormais passer en revue les modèles les plus influents en partant des plus rationalistes pour aller jusqu'à celui considérant le jugement causal comme intrinsèquement biaisé. Nous verrons comment chaque approche intègre les deux critères normatifs du jugement causal et explique l'effet des normes.

Le courant de l'Homme comme avocat

Selon les premières théories de l'attribution causale (Heider, 1958 ; Jones & Davis, 1965 ; Kelley, 1967), l'objectif du sens commun est d'établir des propriétés stables à son environnement social sur la base de comportements observés. Au sein de ces théories, une analogie était couramment faite entre la cognition sociale et la pensée scientifique (e.g., probabilités, statistiques, etc.). Par exemple, selon le modèle de la covariation de Kelley (1967), pour établir des liens causaux, le penseur social prend plusieurs dimensions d'une situation (e.g., sa consistance dans le temps) et calcule un rapport F de manière analogue à celle de la technique statistique de l'analyse de variance (ANOVA). Néanmoins, Kelley (1973) lui-même reconnaîtra que son modèle est « sans doute quelque peu idéalisé » (p. 113).

Comme déjà évoqué, dans les années 80, ces théories ont été vivement critiquées pour leur caractère normatif plutôt que descriptif (e.g., Fincham & Shultz, 1981 ; Hamilton, 1980 ; Shultz et al., 1981, 1986). L'article de Fincham et Jaspars (1980) a constitué un tournant majeur dans la recherche sur le jugement causal¹¹. En s'appuyant sur la philosophie légale de Hart et Honoré (1985)¹² et en l'associant à l'attribution de responsabilité et de blâme pour des événements à l'issue négative (Alicke et al., 2015), Fincham et Jaspars soulignent en effet que ces premiers modèles ne permettraient pas d'expliquer des problèmes complexes de responsabilité. Selon ces auteurs, les situations qui intéressent généralement le sens commun impliquent les notions de responsabilité et de blâme, et se rapprochent plus d'une enquête judiciaire sur un événement spécifique que d'une étude scientifique permettant de trouver des lois universelles. C'est pourquoi ce courant a été nommé le courant de « l'Homme comme avocat » (Fincham & Jaspars, 1980). Cette distinction entre l'étude de l'identification de propriétés stables d'une personne, ou de l'environnement, et celle de savoir si un agent humain a causé ou non un dommage a constitué le principal moteur de ce changement de métaphore du scientifique vers l'avocat (Alicke et al., 2015).

La philosophie légale de Hart et Honoré

L'ouvrage de Hart et Honoré (1985) intitulé « Causation in the Law » a servi d'ancrage théorique aux modèles du courant de « l'Homme comme avocat » et a proposé la première explication du lien entre jugement causal et anormalité. Cet ouvrage est d'ailleurs resté une référence pour des approches plus récentes en psychologie expérimentale (e.g., Hilton et al., 2010, 2016 ; Hilton & Slugoski, 1986 ; McClure et al., 2007). Hart et Honoré (1985) avançaient notamment l'idée que les lois légales, construites par les sociétés depuis des siècles, trouvent

¹¹ Le titre de cet article évoquait explicitement le passage de la conception d'un penseur social « scientifique » à un penseur social « avocat » : « Attribution of responsibility : From Man the scientist to Man as Lawyer »

¹² Les travaux de Hart et Honoré ont été initialement publiés en 1959 dans leur première édition.

leur parallèle dans le jugement causal de sens commun. Cela implique que le jugement causal est considéré comme intrinsèquement conforme aux critères normatifs juridiques. A travers le principe *sine qua non* (i.e., sans l'acte de l'accusé, le dommage ne se serait pas produit), le critère de nécessité est considéré comme le fondement du jugement causal de sens commun.

Néanmoins, les critères de force causale et de causalité réelle n'ont pas été intégrés dans cette analyse. Le seul critère normatif pris en compte par Hart et Honoré en matière de causalité était donc la nécessité, avec une limite que nous avons déjà mise en avant : comment distinguer une cause parmi la multitude de conditions nécessaires à un dommage ? Pour répondre à cette question les deux auteurs avancent que le sens commun, comme le droit, se base sur deux critères de sélection causale. Le premier critère est celui de « l'action humaine volontaire » (p.41). Le second est celui de la condition « anormale par rapport au cours habituel des événements » (p.33). Ils illustrent ces critères par l'exemple suivant : si un bâtiment est détruit par le feu, les conditions nécessaires que sont l'oxygène dans l'air, la présence de matériaux inflammables ou la sécheresse du bâtiment ne seront pas considérées comme des causes par le sens commun car ce sont des conditions « normales ». Le statut de *cause* sera donné à une action humaine volontaire (e.g., un pyromane) ou à une condition anormale (e.g., un court-circuit dans le réseau électrique) (p.11). Évidemment, le caractère anormal d'un événement dépend du contexte, comme illustré par l'exemple suivant :

« Si un incendie se déclare dans un laboratoire ou dans une usine, où des précautions spéciales sont prises pour exclure l'oxygène pendant une partie d'une expérience ou d'un processus de fabrication [...], il ne serait pas du tout absurde dans un tel cas de dire que la présence d'oxygène était la cause de l'incendie. » (p.33).

Toujours selon Hart et Honoré (1985), l'anormalité dépend également de la perspective adoptée : « La cause d'une grande famine en Inde pourra être identifiée par le paysan indien

comme la sécheresse, mais l'Organisation des Nations unies pour l'alimentation et l'agriculture pourra identifier l'échec du gouvernement Indien à faire des réserves comme cause de la famine et la sécheresse comme une simple condition. » (p.33).

La logique derrière cette explication est que lorsqu'un dommage se produit et que nous en demandons la cause, cette demande se base sur la prémisse implicite selon laquelle quelque chose a fait « la différence entre ce qu'il se passe normalement et ce qu'il s'est passé dans cette situation précise » (p.34). Ceci donne l'avantage pratique de se focaliser sur la condition « contrôlable » par les individus permettant de prévenir un futur dommage similaire. C'est pour cette raison que le sens commun préfère désigner un court-circuit (i.e., contrôlable d'un point de vue technique) ou une action humaine volontaire (contrôlable d'un point de vue judiciaire), plutôt qu'une condition non contrôlable telle que la présence d'oxygène dans l'air.

Un jugement séquentiel et rationnel

En s'appuyant sur les travaux de Hart et Honoré, les modèles du courant de « l'Homme comme avocat » ont proposé par la suite une modélisation du jugement causal de sens commun fondée sur des études expérimentales. Ces modèles se focalisent sur la façon dont les individus jugent les agents humains impliqués dans des événements négatifs, et ce, avec toutes les dimensions de ce jugement : jugement causal, jugement de responsabilité, jugement moral et attribution de sanctions. Les participants de ces études étaient ainsi glissés dans le rôle de « jurés fictifs ». A l'instar de Hart et Honoré, ces modèles postulent que l'individu respecte les principes juridiques en procédant de manière séquentielle : jugement causal → jugement de responsabilité → attributions de sanctions (Fincham & Jaspars, 1980 ; Shaver, 1985 ; Shultz et al., 1981). La première étape est donc le jugement causal qui, selon cette perspective, sert uniquement à déterminer si l'agent (donc l'action volontaire ou anormale) est une condition nécessaire au dommage (conformément au principe *sine qua non*). S'il s'avère que ce n'est pas

le cas, le processus de jugement s'achève et la question de la responsabilité et des sanctions ne se pose plus. Si l'agent est une condition nécessaire, alors le jugement de responsabilité aura pour fonction de déterminer si l'acte a été produit de manière volontaire ou négligente. Si c'est le cas, des sanctions peuvent enfin lui être attribuées¹³.

Plusieurs études expérimentales semblent avoir démontré que les individus procèdent selon ces étapes séquentielles (Fincham & Jaspars, 1983 ; Fincham & Roberts, 1985 ; Schleifer et al., 1983 ; Shultz et al., 1981, 1986 ; Shultz & Wright, 1985). Par exemple, Shultz et ses collègues (1981) ont demandé à des participants d'effectuer des jugements de causalité, de responsabilité et de sanctions. Leurs résultats ont montré que le critère de nécessité déterminait le jugement causal et le jugement de responsabilité. De plus, à travers des analyses de trajet, les auteurs concluent que le jugement causal prédit le jugement de responsabilité qui prédit, à son tour, l'attribution de sanctions. Ainsi, les auteurs du courant de « l'Homme comme avocat » ont conclu que les individus sont rationnels car ils incluent le critère normatif de nécessité comme première étape déterminante d'un jugement de culpabilité. Les notions d'anormalité et d'action humaine volontaire, quant à elles, ne sont pas considérées comme des biais, mais comme des outils de discrimination de l'ensemble des conditions nécessaires. En d'autres termes, l'effet de l'infraction d'une norme morale sur le jugement causal est un critère additionnel permettant de compléter un modèle normatif incomplet qui n'inclut que la nécessité réelle.

Limites du courant de « l'Homme comme avocat »

Le courant de « l'Homme comme avocat » a constitué un tournant majeur dans la recherche sur l'attribution causale en se focalisant sur les jugements causaux vis-à-vis

¹³ Il convient toutefois de noter que cette conception d'un jugement de responsabilité débutant par une analyse rationnelle du lien causal en termes de nécessité pour ensuite évaluer les états mentaux de l'individu se retrouve déjà chez Heider (1958, p.113).

d'événements négatifs et spécifiques. A travers les travaux de Hart et Honoré, ce courant a également permis d'apporter la première explication de l'effet de l'anormalité sur le jugement causal. Néanmoins, cette explication comprend plusieurs limites et a reçu de nombreuses critiques (pour une revue, voir Stapleton, 1988 ; Summers, 2018).

Premièrement, la méthodologie de Hart et Honoré est discutable. En effet, ces auteurs ont déterminé les principes du jugement causal de sens commun en analysant la façon dont les individus parlent de causalité dans le langage ordinaire. La prémisse selon laquelle le langage ordinaire est la clé de clarification des difficultés conceptuelles de la causalité était courante dans les années 50. Elle est cependant totalement abandonnée de nos jours (Summers, 2018). La théorie de Hart et Honoré a ainsi été établie davantage par « affirmation que par preuves » (Cane, 2002, p.129). Par ailleurs, il s'agit du langage ordinaire tel que ces deux auteurs l'imaginaient, et non de l'étude d'une base objective de données linguistiques. Par conséquent, même si des études ont testé ces hypothèses par la suite, la théorie de Hart et Honoré ne repose pas sur une approche scientifique du jugement causal.

Deuxièmement, bien que ce jugement causal soit conçu comme la première étape d'un processus censé aboutir à l'attribution d'une responsabilité et d'éventuelles sanctions, un seul critère normatif permet de le qualifier : la nécessité des actions de l'agent pour la survenue du dommage (ou principe *sine qua non*). Comme nous l'avons vu, ce critère n'est pas suffisamment discriminant face à la multitude de conditions nécessaires à la survenue du plus simple des événements. Selon cette approche, seuls des critères de sélection causale, tels que l'anormalité et les actions humaines volontaires, permettent de discriminer une cause parmi l'ensemble des conditions nécessaires. Or, ces critères ne sont pas des critères normatifs pour le jugement causal. En somme, la notion même de jugement rationnel et conforme aux principes juridiques peut être discutée car le modèle normatif considéré par ce courant est

incomplet (i.e., non prise en compte de la notion de force causale permettant de discriminer les conditions nécessaires entre elles).

Troisièmement, le processus psychologique sous-jacent au jugement causal reste peu précis. Aucun auteur de ce courant n'explique réellement comment les individus procèdent pour déterminer qu'un facteur est une condition nécessaire au dommage. Des modèles plus récents sur le jugement moral (e.g., Cushman, 2008 ; Malle et al., 2014) s'inscrivent toujours dans cette conception d'un processus séquentiel rationnel, allant du jugement causal jusqu'au blâme et aux sanctions. Comme les modèles du courant de « l'Homme comme avocat », ces modèles ne détaillent pas les processus psychologiques impliqués dans le jugement causal. Cette première étape qu'est le jugement causal est considérée par défaut comme rationnelle et conforme au critère de nécessité. Pourtant, plusieurs travaux que nous détaillerons un peu plus loin ont remis en question cet ordre séquentiel ainsi que la rationalité du jugement causal (e.g., Alicke, 1992 ; Alicke et al., 2011 ; Lagnado & Channon, 2008 ; Rogers et al., 2019).

Les limites que nous venons de décrire n'ayant pas été accommodées par le courant de « l'Homme comme Avocat », d'autres approches ont été proposées. Parmi elles, un courant dominant à l'heure actuelle est celui des modèles du raisonnement contrefactuel. Nous présenterons ces derniers dans la prochaine section.

L'approche du raisonnement contrefactuel

*« Without considering alternatives to reality, we must accept the past as having been inevitable and must believe that the future will be no different from the past. »
Johnson et Sherman (1990, p. 150)*

Les modèles du raisonnement contrefactuel ont pris la suite des modèles du courant de « l'Homme comme Avocat » dans l'explication des jugements causaux de sens commun. Ces modèles permettent non seulement une explication du processus cognitif permettant de détecter

la nécessité réelle et la sensibilité des conditions, mais aussi une explication de l'effet des normes.

Déterminer la nécessité réelle

La rationalité et l'imagination ont souvent été considérées comme antagonistes. Pourtant, elles sont étroitement liées dans le domaine du jugement causal (Byrne, 2007). Selon les modèles du raisonnement contrefactuel, pour déterminer la présence d'un lien causal, nous devons imaginer des situations ne s'étant pas produites, c'est-à-dire effectuer des simulations mentales (Byrne, 2002, 2016 ; Byrne & Johnson-Laird, 2020 ; Goldvarg & Johnson-Laird, 2001 ; Kahneman & Tversky, 1982 ; Mandel, 2011 ; Mandel et al., 2005 ; Van Hoek et al., 2015). Plus précisément, comme nous l'avons vu avec l'approche philosophique sur le critère de nécessité réelle, pour déterminer un lien causal, deux situations doivent être contrastées : une situation dans laquelle la cause supposée est présente et une situation dans laquelle elle ne l'est pas (e.g., Lewis, 1973, 2000 ; Mackie, 1980). Ce n'est que de cette façon que nous pouvons conclure si l'effet étudié se serait produit en l'absence de la cause supposée. Or, si les scientifiques réussissent cette opération grâce à l'utilisation d'un groupe expérimental (cause supposée présente) et d'un groupe contrôle (cause supposée absente), le sens commun, quant à lui, n'expérimente qu'une seule situation à la fois. Il s'agit généralement d'une situation dans laquelle la cause supposée est présente : l'accusé a tué la victime, la cigarette a enflammé la forêt, le chat a cassé le vase, etc. Pour déterminer si la cause supposée était une condition nécessaire à l'effet étudié, nous allons devoir effectuer une simulation mentale d'une situation dans laquelle la cause supposée ne se produit pas. Cette simulation est dite *contrefactuelle* puisqu'il s'agit d'un état possible du monde ne s'étant pas effectivement produit. Pour produire un jugement causal, nous devons observer, au sein de cette simulation, si l'effet se produit toujours lorsque la cause supposée est mentalement niée. Si l'effet disparaît, nous concluons que la cause supposée était effectivement une cause de cet effet. En revanche, si l'effet persiste,

nous concluons que la cause supposée n'en était pas une. Par exemple, si le campeur n'avait pas laissé tomber une allumette, la forêt n'aurait pas brûlé ; si l'étudiant n'avait pas écrit une rédaction intéressante, il n'aurait pas été accepté en Master ; si l'accusé n'avait pas tiré sur la victime, celle-ci ne serait pas décédée, etc. De plus, ce contraste permet de distinguer la causalité d'une simple corrélation¹⁴ (Fenker et al., 2005 ; Lagnado, 2011 ; Satpute et al., 2005). Ainsi, la simulation contrefactuelle nous permet d'aller au-delà de l'observation de successions d'événements ou de simples corrélations en distinguant les « véritables » liens de causalité.

Depuis les travaux pionniers de Kahneman et Tversky (1982) sur le sujet, de nombreux travaux récents en psychologie cognitive¹⁵ (e.g., Gerstenberg et al., 2021 ; Henne et al., 2021 ; Kominsky et al., 2015 ; Kominsky & Phillips, 2019) ont démontré la place centrale de la pensée contrefactuelle dans le jugement causal face à des événements purement physiques (e.g., des boules de billard en mouvement), ou des événements sociaux de faible gravité (e.g., une alarme qui se déclenche). Par exemple, une étude d'oculométrie (Gerstenberg et al., 2017) apporte un soutien empirique à l'hypothèse des simulations contrefactuelles pour déterminer si une cause supposée est une condition nécessaire réelle. Dans cette étude, les participants devaient regarder une vidéo dans laquelle des boules de billard s'entrechoquent sur une scène comportant des murs et une porte (voir Figure 3, p.42). Dans la vidéo, la balle B se dirige vers la porte à gauche (en rouge) mais la balle A surgit du coin inférieur droit et percute la balle B. Ce choc modifie la trajectoire de la balle B et l'empêche de sortir de la pièce. Les résultats ont montré que lorsque les participants doivent déterminer si la balle A a empêché la balle B de passer la porte, ils produisent des saccades « contrefactuelles » (représentés par les points blancs sur la

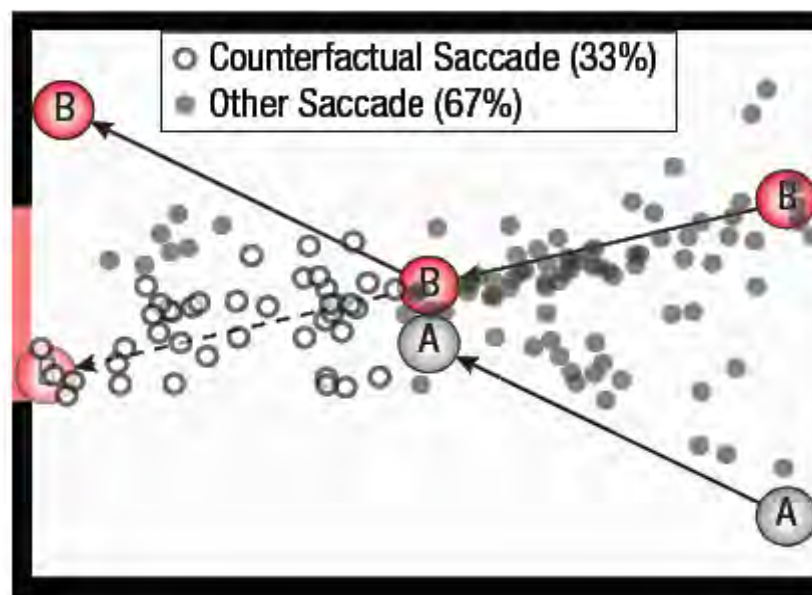
¹⁴ Imaginons qu'un individu soit face à une simple corrélation telle que les ventes de glace et les coups de soleil. Dans ce cas, imaginer une situation dans laquelle les ventes de glace sont niées n'empêche pas la présence des coups de soleil de survenir puisque la véritable cause est la période de l'année durant laquelle les deux événements surviennent.

¹⁵ Bien que le jugement causal soit une compétence cognitive fondamentale pour l'être humain, l'étude de ce processus a été longtemps réservée à la psychologie sociale et n'a été réellement intégré à la recherche en psychologie cognitive que dans les années 90 (Waldmann & Hagmayer, 2013).

Figure 3). Plus précisément, ces saccades oculaires suivent le trajet que la balle B aurait eu si la balle A n'avait pas été présente dans la scène. Les auteurs concluent que les participants utilisent leur compréhension intuitive de la physique pour simuler mentalement ce qui se serait passé si la cause candidate (i.e., la balle A) n'était pas intervenue.

Figure 3.

Résultats des saccades oculaires des participants de l'étude de Gerstenberg et al (2017).



Note. Reproduit depuis *Eye-Tracking Causality* (p. [8], fig. 3), par T. Gerstenberg et al., 2014, *Psychological Science*. <https://doi.org/10.1177/095679761771305>. Reproduit avec permission.

D'autres études empiriques soutiennent la conception contrefactuelle du jugement causal en montrant que les individus considèrent également les omissions (e.g., ne pas arroser une plante) comme des causes (Henne et al., 2017 ; Livengood & Machery, 2007). La recherche sur le lien entre simulation contrefactuelle et jugement causal s'est également étendue aux situations impliquant des agents humains pour des situations bénignes (e.g., Kominsky et al., 2015 ; Kominsky & Phillips, 2019), et a même étendu ses résultats au-delà des populations

WEIRD¹⁶ (Le Guen et al., 2015 ; Wente et al., 2022). Ainsi, cette approche soutient l'idée que les individus sont capables, à travers des simulations contrefactuelles, d'évaluer si une cause supposée est une condition nécessaire réelle à un résultat. De plus, ces simulations joueraient un rôle majeur et déterminant dans le jugement causal de sens commun.

Par conséquent, ce jugement causal serait parfaitement rationnel vis-à-vis du critère normatif le plus important en matière de causalité, à savoir la nécessité réelle. De nombreux modèles sur la cognition humaine et sur le raisonnement proposent cependant que nous n'évaluons pas notre environnement de façon dichotomique mais de façon continue (e.g., Griffiths et al., 2012 ; Sanborn et al., 2010 ; Spellman & Gilbert, 2014). Il en irait de même pour le jugement causal (Gerstenberg et al., 2021 ; Lagnado et al., 2013 ; Lagnado & Gerstenberg, 2017 ; Stephan & Waldmann, 2018).

Déterminer la sensibilité

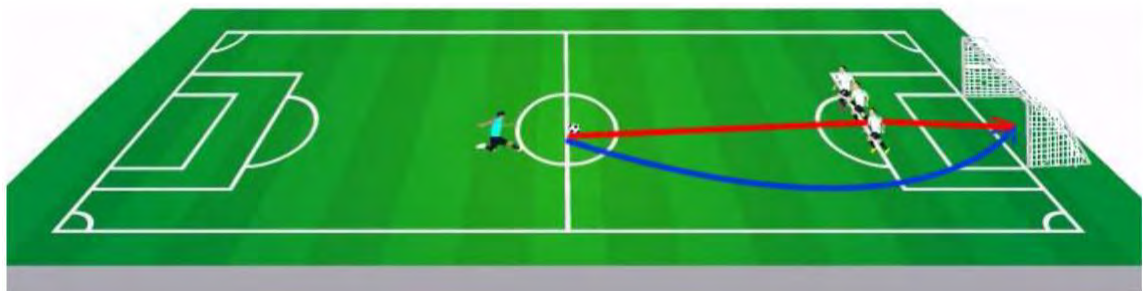
Les premières théories du jugement causal intégrant la notion de force causale l'associaient à la notion de probabilité (e.g., Cheng & Novick, 1990 ; Hitchcock, 1997 ; Spellman, 1997). Selon ces théories, la force de la relation entre une cause et son effet correspond à la mesure avec laquelle la cause augmente la probabilité d'apparition de l'effet. Néanmoins, plusieurs travaux ont remis en question l'utilisation des probabilités dans le jugement causal de sens commun (Grinfeld et al., 2020 ; Hilton et al., 2010 ; Mandel, 2003 ; McClure et al., 2007). Des modèles plus récents (Gerstenberg et al., 2020 ; Grinfeld et al., 2020 ; Icard et al., 2017 ; Kominsky et al., 2015) ont proposé que les simulations contrefactuelles permettaient d'évaluer la force causale en plus de la nécessité réelle. Le principe reste le même que pour la nécessité réelle, mais cette fois-ci la simulation contrefactuelle se focalise sur les conditions adjacentes à la cause supposée. Si la négation de ces conditions annule le lien entre

¹⁶ Western, educated, industrialized and democratic (Occidental, éduqué, industrialisé, riche et démocratique).

la cause supposée et le résultat, alors cette cause supposée est jugée sensible (i.e., faible lien causal). A l'inverse, si la négation de ces conditions n'annule pas le lien causal, celui-ci est jugé insensible (i.e., fort lien causal). Dans l'étude de Grinfeld et ses collègues (2020), les participants devaient regarder une vidéo dans laquelle un joueur de foot tirait au but face à un mur de défenseurs en mouvement (voir Figure 4 ci-dessous).

Figure 4.

Illustration de tirs au but sensible (ligne rouge) et insensible (ligne bleue).



Note. Reproduit depuis *Causal Responsibility and Robust Causation* (p. [8], fig. 3), par G. Grinfeld et al., 2020, *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2020.01069>. Reproduit avec permission. Vidéos disponibles : <https://github.com/guygrinfeld/Responsibility-and-Robust-Causation-Experiments/tree/master/videos-sample>

La sensibilité du lien entre le tir du joueur et le but variait. Dans une condition, le tir passait par le mur de défenseurs (ligne rouge dans la Figure 4). Dans l'autre condition, la balle le contournait (ligne bleue dans la Figure 4). Dans les deux cas, le tir du joueur est une condition nécessaire réelle au but. Néanmoins, les participants attribuaient un score de causalité plus élevé au joueur lorsque la balle contournait le mur de défenseurs. Ceci s'explique par le fait que le lien causal du tir en rouge dépend du mouvement des défenseurs. Si le mouvement des défenseurs avait été différent, la balle aurait pu être stoppée. Ceci n'est évidemment pas valable

lorsque la balle contourne les défenseurs. Le tir en bleu possède donc une force causale plus élevée (i.e., insensibilité aux défenseurs) que le tir en rouge (i.e., sensibilité aux défenseurs).

Cette notion de sensibilité permet d'expliquer un grand nombre de résultats ayant démontré qu'un agent est davantage considéré comme la cause d'un effet qu'il a directement provoqué, plutôt que la cause d'un effet qui résulte d'une longue chaîne causale (Brickman et al., 1975 ; Hilton et al., 2010 ; Lagnado & Channon, 2008 ; McClure et al., 2007 ; Spellman, 1997). En effet, plus la chaîne causale entre les actions d'un agent et un dommage est longue, plus le lien causal les unissant dépend des conditions intermédiaires présentes dans la chaîne. En outre, le jugement du lien causal entre une cause et son effet dépend en partie du statut des autres conditions (Kominsky et al., 2015 ; McGill & Tenbrunsel, 2000). A nouveau, le jugement causal de sens commun, en captant la force causale des relations de son environnement, serait donc rationnel.

Un jugement causal rationnel ?

Comme nous venons de le voir, le jugement causal de sens commun pourrait donc ici être considéré comme rationnel. D'un point de vue adaptatif, ce minimum de rationalité n'est pas surprenant. Savoir déterminer correctement la présence et la force d'un lien causal est primordial pour comprendre le monde qui nous entoure et prévoir ce qui peut advenir (Kuhn, 2012). Savoir que les coups de soleil sont provoqués par les rayons du soleil plutôt que par les ventes de glaces permet d'éviter d'en recevoir. Ces simulations contrefactuelles rationnelles seraient réalisées automatiquement¹⁷(Goldinger et al., 2003 ; Van Hoek et al., 2015). Ceci vient à l'encontre de l'idée selon laquelle les « bonnes réponses » (i.e., application des critères normatifs) seraient liés à un système de pensée plus lent, plus réflexif et plus coûteux en

¹⁷ Lorsque nous parlerons de processus automatiques dans cette thèse, nous ferons référence à un processus rapide et peu coûteux en mémoire de travail (Evans & Stanovich, 2013).

mémoire de travail (pour une revue des théories duales du raisonnement, voir, e.g., De Neys, 2018). Cet aspect automatique des simulations contrefactuelles s'inscrit en revanche dans la lignée de travaux récents démontrant que les heuristiques de jugement peuvent être parfaitement logiques et rationnelles (Bago & De Neys, 2017, 2019), à condition d'avoir acquis les connaissances pertinentes pour la tâche (De Neys, 2012 ; Stanovich, 2018). Concernant la simulation contrefactuelle, celle-ci ne nécessite pas de connaissances détaillées pour être produite, de simples connaissances intuitives et élémentaires sur le monde suffisent (Lagnado, 2011 ; Lagnado et al., 2013). Prenons l'exemple d'un accusé qui empoisonne une victime. Pour déterminer la culpabilité de l'accusé, le jury va notamment effectuer une simulation mentale d'une situation dans laquelle l'accusé ne verse pas de poison. Une connaissance précise du processus biologique allant de l'ingestion de la substance toxique à la privation d'oxygène du cerveau n'est pas requise pour cette simulation. De simples connaissances générales selon lesquelles l'ingestion d'un poison provoque la mort sont suffisantes.

En résumé, les nombreux travaux de cette approche montrent que les individus seraient capables de déterminer automatiquement le modèle causal d'une situation de manière rationnelle. Des travaux ont également étendu le raisonnement contrefactuel comme déterminant pour le jugement de responsabilité (Grinfeld et al., 2020 ; Lagnado et al., 2013). Selon cette approche, celui-ci serait également rationnel. Néanmoins, l'effet des normes reste toujours à expliquer et nous verrons comment ces modèles du raisonnement contrefactuelles l'ont intégré.

La focalisation sur la condition anormale

Selon le courant du raisonnement contrefactuel, l'effet des normes trouve son origine dans la tendance des individus à effectuer des simulations contrefactuelles vis-à-vis d'événements inattendus et surprenants (Dyczewski & Markman, 2012 ; Hastie, 1984 ; Kahneman & Tversky, 1982 ; Kanazawa, 1992 ; Weiner, 1985). L'attention portée aux

CHAPITRE 2

DE L'EFFET DES NORMES À LA VALIDATION DU BLÂME

événements lorsque nos attentes sont contredites est, en soi, la manifestation d'une pensée contrefactuelle : « normalement, cela n'aurait pas dû se passer comme cela » (Mandel, 2005). En d'autres termes, ces simulations s'effectuent sur des situations qui dévient de notre représentation du cours normal des événements (Hilton & Slugoski, 1986 ; Kahneman & Miller, 1986 ; Phillips et al., 2015). Les modèles du raisonnement contrefactuel ne se focalisent pas uniquement sur les normes morales (Kominsky & Phillips, 2019). Ces normes peuvent tout aussi bien être statistiques (e.g., du soleil en été), prescrites (e.g. respecter les limitations de vitesse) que morales¹⁸ (e.g., ne pas tricher). La première étude sur la simulation contrefactuelle (Kahneman & Tversky, 1982) illustre parfaitement cette considération des normes. Les participants étaient invités à lire un scénario dans lequel Monsieur Jones quittait le bureau à son heure habituelle, mais rentrait chez lui par un chemin inhabituel et se faisait tuer par un camion qui le percutait. Après avoir lu le texte, les participants devaient indiquer comment les choses auraient pu se passer autrement pour éviter la mort de Monsieur Jones. Les résultats ont montré que les participants avaient tendance à imaginer que Monsieur Jones ne serait pas mort s'il était rentré par le chemin habituel. Dans une autre condition, Monsieur Jones rentrait par le chemin habituel mais partait plus tôt que d'habitude. Dans ce cas, les participants avaient davantage tendance à imaginer un scénario dans lequel Monsieur Jones partait à son heure habituelle. Ces résultats montreraient que les individus ont tendance à rétablir le cours normal des événements lorsqu'ils doivent imaginer des alternatives (i.e., effectuer des simulations contrefactuelles). Les modèles du raisonnement contrefactuel suggèrent que l'objectif principal des jugements causaux est d'identifier le meilleur point d'intervention, c'est-à-dire le facteur sur lequel nous devons intervenir pour empêcher des événements négatifs similaires à l'avenir

¹⁸ Les normes morales et les normes prescrites sont souvent assimilées mais elles peuvent parfois se distinguer : par exemple, la ségrégation raciale aux Etats-Unis était une norme prescrite par la loi mais certaines personnes y voyaient l'infraction d'une norme morale.

(Willemsen & Kirfel, 2019). Rétablir le cours normal et attendu des événements permet ainsi d'éviter de nouveaux événements négatifs et imprévus.

En s'appuyant sur les modèles probabilistes de la cognition (e.g., Chater et al., 2006 ; Griffiths et al., 2012 ; Knill & Pouget, 2004 ; Vilares & Kording, 2011 ; Yang & Shadlen, 2007), Icard et ses collègues (2017) proposent que certaines simulations contrefactuelles ont une plus grande probabilité d'activation que d'autres. Ceci s'explique par le fait que nous n'avons pas la capacité de générer un nombre illimité de simulations. La probabilité d'activation d'une simulation sera déterminée par la normalité (statistique, prescrite ou morale) des événements. En effet, face à un événement anormal, les individus formulent davantage de propositions contrefactuelles qu'ils considèrent comme statistiquement fréquentes ou moralement bonnes (Kahneman & Miller, 1986 ; McCloy & Byrne, 2000 ; Phillips et al., 2015). En revanche, lorsque l'action d'un agent est normale, les possibilités contrefactuelles impliquant une action alternative anormale ne viennent pas si naturellement à l'esprit (Petrocelli et al., 2011). Par exemple, si un élève échoue parce qu'il n'a pas révisé, il est plus facile de générer une simulation contrefactuelle dans laquelle l'élève révisé (i.e., simulation contrefactuelle normale), plutôt que d'imaginer que son instituteur a été enlevé par des extra-terrestres (i.e., simulation contrefactuelle anormale).

En d'autres termes, les simulations contrefactuelles produites dans une tâche de jugement causal sont une version « normalisée » de ce qui s'est produit de manière inattendue et surprenante. Le critère de normalité apporte alors un cadre limitant le nombre de simulations contrefactuelles, ainsi qu'un guide efficace pour déterminer le meilleur point d'intervention (Hitchcock & Knobe, 2009 ; Phillips et al., 2019 ; Phillips & Cushman, 2017). Si nous reprenons l'exemple de la *pen vignette* (cf. description p.31-32), le point d'intervention le plus pertinent est le professeur Smith et non l'assistant administratif. D'un point de vue pratique, si

CHAPITRE 2 DE L'EFFET DES NORMES À LA VALIDATION DU BLÂME

nous voulons éviter une nouvelle pénurie de stylos, il convient de rappeler à l'ordre le professeur sur le règlement de l'Université (i.e., la norme prescrite). S'assurer que la norme prescrite est respectée et que les événements vont continuer à suivre leur cours normal est une stratégie efficace pour prévenir de nouveaux événements négatifs tels que les pénuries de stylos. Cette stratégie serait ainsi rationnelle, au sens d'une *rationalité limitée* ou *écologique*, puisqu'elle permettrait de n'utiliser nos ressources cognitives limitées que pour des interventions pertinentes sur notre environnement. Néanmoins, rappelons que nous avons choisi de suivre, comme la majorité des auteurs, la définition de la *rationalité normative* (i.e., le respect de critères normatifs, ici la *nécessité réelle* et la *sensibilité*). L'effet des normes constitue ici un biais de jugement puisqu'il s'agit d'une déviation par rapport à l'utilisation exclusive des critères normatifs (nécessité réelle et sensibilité). Toutefois, nous verrons que l'écart par rapport aux critères normatifs peut être plus ou moins important. Dans le cas de l'approche contrefactuelle, ce biais est relativement limité puisque l'application des critères de nécessité réelle et de sensibilité n'est pas affectée par l'anormalité. Il s'agit seulement d'un critère qui s'additionne pour des raisons pragmatiques¹⁹ liées à l'objectif principal du jugement causal qui est de contrôler l'environnement. Dans l'approche du raisonnement biaisé portée par le Modèle du Contrôle Coupable, comme nous le verrons dans la prochaine partie, les choses sont bien différentes car l'infraction d'une norme morale impacte directement l'évaluation des critères normatifs.

Pour résumer, nos connaissances sur le monde nous permettent de faire des simulations contrefactuelles. Ces simulations nous permettent d'évaluer la nécessité réelle et la sensibilité des conditions des événements observés. Les infractions aux normes, quant à elles, constitueraient un point d'intervention efficace sur lequel les simulations contrefactuelles

¹⁹ Le terme *pragmatique* est ici employé pour désigner ce qui concerne la pratique. Ce terme sera utilisé avec un autre sens dans le dernier chapitre.

opèrent. A l'instar du courant de « l'Homme comme avocat », les jugements causaux sont donc plutôt rationnels puisqu'ils intègrent les critères normatifs dans les processus cognitifs du jugement causal. En revanche, ces jugements se basent sur un facteur extra-normatif (i.e., les normes) pour accomplir leur objectif principal : prévenir les événements négatifs et reproduire les événements positifs. Bien que cette approche repose sur de nombreux arguments théoriques et empiriques, celle-ci présente certaines limites (voir Collins et al., 2004 ; Moore, 2009 ; Willemsen & Kirfel, 2019). Nous allons présenter trois de ces limites qui nous permettront par la suite d'introduire l'approche du biais.

Limites de l'approche contrefactuelle

Une première limite a été soulevée par Mandel (2003). Cet auteur propose que les jugements causaux et les jugements contrefactuels diffèrent dans certains cas, notamment les cas de préemption (i.e., un facteur aurait causé l'effet si un autre ne l'avait pas causé avant). Dans son étude, les participants devaient lire un récit dans lequel Monsieur Wallas, une figure importante du crime organisé, se fait empoisonner par un rival mais décède après avoir été percuté par la camionnette d'un autre rival avant que le poison ne fasse effet. Après avoir lu le texte, les participants devaient répondre à une question de jugement causal (« Veuillez énumérer jusqu'à quatre facteurs que vous considérez comme des causes de la mort de M. Wallace. ») et à une question de raisonnement contrefactuel (« Veuillez énumérer jusqu'à quatre façons dont l'histoire pourrait être modifiée de sorte que le résultat soit différent. »). Les résultats ont montré que les participants citaient principalement la camionnette pour la question de jugement causal, mais ils citaient principalement la vie de criminelle de Monsieur Wallas à la question de raisonnement contrefactuel. Mandel propose ici que le jugement causal se focalise sur la cause réelle directe du dommage (i.e., l'accident), quand le raisonnement contrefactuel permet de détecter le facteur permettant de prévenir à la fois la mort réelle de Monsieur Wallas (i.e. l'accident) et une mort similaire hautement probable dans un univers

contrefactuel (i.e., l'empoisonnement). Cette première limite représentait la dissociation pouvant exister entre la pensée contrefactuelle et le jugement causal (pour d'autres études depuis, Hilton et al. 2021 ; Lombrozo, 2010 ; Walsh & Sloman, 2011).

La deuxième limite porte directement atteinte à la conception rationnelle de l'approche contrefactuelle. Si le jugement causal est aussi rationnel que le laisse supposer cette approche, comment expliquer l'effet de facteurs extra-légaux sur le jugement causal ? En effet, de nombreuses études ont montré l'effet de nombreux facteurs extra-normatif sur les jugements causaux, tels que la valence du résultat (Alicke et al., 2011 ; Reuter et al., 2014), la personnalité de l'agent (Alicke & Zell, 2009 ; Nadler, 2012), ou encore l'attractivité de la victime (Alicke et al., 2008, 2011). Des résultats similaires ont été observés sur les attributions de responsabilité avec une influence de la sévérité du dommage (Burger, 1981 ; Robbennolt, 2000), de l'attractivité physique de l'accusé (Efran, 1974 ; Saladin et al., 1988), de l'ethnie du jury (Anwar et al., 2012), ou encore de l'apparence de la victime (Janoff-Bulman et al., 1985). La conception rationaliste de l'approche contrefactuelle, selon laquelle les jugements causaux et de responsabilité respectent les critères normatifs et utilisent uniquement les normes comme point d'intervention, ne permet pas de rendre compte de ces résultats.

Une troisième limite est que la majorité des travaux récents démontrant la capacité des simulations contrefactuelles à prendre en compte les critères normatifs de nécessité réelle et de sensibilité (Gerstenberg et al., 2017, 2021 ; Grinfeld et al., 2020) ne l'ont testé que sur des événements physiques (e.g., boules de billard, dés lancés, ballons de foot). Gerstenberg et ses collègues (2021) vont même jusqu'à intituler leur modèle : « Un modèle de simulation contrefactuelle des jugements causaux pour les événements physiques ». Des travaux mettent en avant deux raisons de dissocier les jugements vis-à-vis d'agents humains de ceux vis-à-vis d'objets physiques (Newman et al., 2010 ; Strickland et al., 2017 ; Van Overwalle, 2011 ;

Walsh & Byrne, 2007). Premièrement, les connaissances préalables requises pour effectuer une simulation contrefactuelle ne sont pas les mêmes. Pour comprendre les actions d'un agent humain, des connaissances en théorie de l'esprit sont nécessaires (Saxe et al., 2005), alors que pour les événements physiques, ce sont des connaissances en physique ou en biologie qui sont requises (Johnson & Ahn, 2015). Cette idée avait déjà été évoquée par Buss (1978) en distinguant les « raisons » et les « causes ». Selon cet auteur, les explications pour un comportement humain ne peuvent pas se contenter de décrire les mouvements du corps de la personne (comme nous le ferions pour un objet physique). Pour expliquer un comportement humain, il faut notamment mobiliser la notion de raison (i.e., dans quel but l'agent a-t-il produit ce comportement ? l'a-t-il fait intentionnellement ?) afin de rendre l'action humaine intelligible. Lorsqu'une tâche de jugement causal implique des agents humains, s'ajoutent des notions de normes morales, d'intentions et d'actions volontaires. Par conséquent, si les participants respectent rigoureusement les critères normatifs du jugement causal pour des tâches de jugement envers des objets physiques, cela n'implique pas nécessairement que ce jugement ne soit pas biaisé lorsque des agents humains sont impliqués. Deuxièmement, il semblerait que la majorité des individus ait une représentation dualiste de la causalité, c'est-à-dire de considérer la cognition humaine comme séparée du monde physique (Bloom, 2006 ; Heider, 1958 ; Strickland et al., 2017). Ceci s'expliquerait par une tendance naturelle de l'individu à croire au libre arbitre. Pour Bloom (2006) : « Les gens pensent universellement que la conscience humaine est séparée du monde physique [...]. Et presque tout le monde croit au libre arbitre. Tant au niveau phénoménologique qu'au niveau intellectuel, nous nous considérons comme des agents libres. Même si notre corps est physique et peut être affecté par des choses physiques, nous avons le choix. » (p.211-212). Dit autrement, les objets physiques sont déterminés par les lois du monde, alors que le comportement des individus est régi par un

libre arbitre n'obéissant qu'à ses propres lois. Nous approfondirons la question de la spécificité des jugements vis-à-vis d'événements sociaux dans la prochaine partie.

Conclusion

A l'instar des scientifiques, les profanes cherchent à être corrects lorsqu'ils effectuent un jugement causal. Par conséquent, ce jugement causal est généralement associé à une « force normative » (Waldmann, 2017, p.2), c'est-à-dire une volonté de comprendre le monde tel qu'il est afin de s'y adapter efficacement. Par exemple, l'identification correcte de la cause des difficultés d'un enfant dans ses études (dues par exemple à de la dyslexie plutôt qu'à des problèmes émotionnels) permettra de prendre les mesures correctives appropriées (Hilton, 2002, p.218). A travers la notion de simulation contrefactuelle, nous avons présenté la capacité du sens commun à utiliser les critères normatifs de nécessité réelle et de sensibilité. Par conséquent, l'individu serait rationnel lorsqu'il doit juger les liens de causalité de son environnement. L'effet des normes, quant à lui, constituerait un outil permettant de sélectionner la condition la plus pertinente sur laquelle cette simulation contrefactuelle doit opérer. Néanmoins, malgré la volonté du sens commun de comprendre correctement son environnement, il est probable qu'il existe un écart entre les théories normatives et descriptives du jugement causal. En effet, cette approche ne permet pas d'expliquer les résultats qui dévient du cadre normatif. Nous verrons dans la prochaine partie que la prise en compte de la dimension motivée du jugement causal concernant l'effet des normes pourrait expliquer ces déviations.

La dimension motivée du jugement causal

« Behavioral observers are not only rational calculators but also evaluators whose sympathies and antipathies affect their calculations. »

Rogers et al. (2019, p.2)

La notion de *raisonnement motivé* a été développée pour illustrer l'idée selon laquelle l'atteinte de la « vérité » n'est pas le seul objectif des raisonneurs (Kunda, 1990). D'autres objectifs entrent en jeu, tels que le maintien d'une estime de soi positive (e.g., Valdesolo & DeSteno, 2008), la conservation de nos croyances (e.g., Lord et al., 1979), la diminution de l'anxiété (e.g., Ditto et al., 2003), ou encore la volonté de punir un individu blâmable (e.g., Alicke, 1992). Autrement dit, nous sommes parfois davantage motivés à parvenir à des conclusions souhaitées qu'à des conclusions correctes. Dans le domaine du jugement causal, le principal modèle à avoir placé cette dimension motivée au cœur du jugement de sens commun est le Modèle du Contrôle Coupable (CCM, Alicke, 2000).

Selon le CCM, si dans la *pen vignette*, le Professeur Smith est davantage désigné comme cause de la pénurie de stylos, ce n'est pas simplement lié à son infraction d'une norme prescrite. En réalité, le jugement causal de sens commun serait ici biaisé par les évaluations spontanées négatives suscitées par ce « voleur de stylos malintentionné » (Alicke et al., 2011, p.31). L'objectif de blâmer ce Professeur « malveillant » viendrait ainsi interférer avec l'objectif de comprendre rationnellement les liens causaux de la situation. Or, dans les situations qui nous intéressent (i.e., situations à l'issue négative impliquant un agent humain), les évaluations négatives et les jugements de blâme sont omniprésents. Les observateurs sociaux cherchent, certes, à comprendre ce qu'il s'est passé, mais ils vont également tenter de détecter si quelqu'un doit être blâmé et sanctionné pour le dommage survenu.

Avant d'entrer plus en détails sur les explications du CCM concernant l'effet des normes, nous allons nous intéresser à cet objectif d'ordre moral. Dans un premier temps, nous verrons que l'évaluation des états mentaux de nos semblables occupe une place centrale dans la cognition humaine. Nous verrons ensuite que lorsque nous détectons des états mentaux pernicioeux, nous sommes affectés par des évaluations spontanées négatives. Enfin, nous détaillerons les propositions du CCM selon lesquelles ces évaluations négatives peuvent conduire à un jugement causal biaisé.

L'évaluation des états mentaux de nos semblables

A travers l'approche du raisonnement contrefactuel, nous avons vu l'importance pour l'individu de comprendre son environnement physique. Nous allons désormais détailler l'importance d'évaluer et de maîtriser son environnement social. Savoir discriminer qui est un partenaire social approprié de qui ne l'est pas serait aussi indispensable pour l'individu que de comprendre les liens de causalité de son environnement (Baumard et al., 2013 ; Debove et al., 2015 ; Geraci & Surian, 2011), et ce, quelle que soit la culture d'appartenance des individus (Curry et al., 2019). Ainsi, la capacité à distinguer les « méchants » des « gentils » apparaîtrait de façon très précoce (Bloom, 2013 ; Hamlin, 2013 ; Holvoet et al., 2016). Par exemple, dès l'âge de 6 mois, les nourrissons préfèrent un individu qui en aide un autre à un individu neutre, et préfèrent un individu neutre à un individu malveillant (Hamlin et al., 2007). Cette compétence évaluative est tellement fondamentale, qu'en réalité, la quasi-totalité des actions humaines significatives seraient automatiquement évaluées (Alicke et al., 2011 ; Malle et al., 2014). Or, sur quoi repose cette évaluation ? Comment déterminer si un individu est un « mauvais » partenaire social, voire « nuisible » à la vie en communauté ? Cette dimension évaluative repose principalement sur la cognition sociale, renvoyant notamment à l'estimation des états mentaux de nos semblables (Alicke, 2000 ; Cushman, 2008 ; Guglielmo et al., 2009).

En psychologie, l'étude de la cognition sociale a mis en avant une tendance importante de l'être humain qui consiste à surestimer la présence d'intentions dans son environnement (e.g., Rosset, 2008). Nous pouvons notamment citer la célèbre étude de Heider et Simmel (1944), qui a montré que les participants attribuent des intentions à des formes géométriques en mouvement (un grand triangle, un petit triangle et un cercle, voir Figure 5, p.57)²⁰. Dans cette étude, 49 participants sur 50 utilisaient le langage de l'intentionnalité pour décrire ces formes élémentaires (e.g., « chasser », « fuir », « suivre »). Les travaux en neurosciences de Castelli et ses collègues (2000) ont montré que, face à ces formes géométriques, le sillon temporal supérieur s'activait. Or, cette zone est directement impliquée dans la perception des intentions d'autrui. Malle et Holbrook (2012) ont voulu, quant à eux, comparer la durée d'activation de différentes inférences sociales : intentionnalité, désirs, croyances et personnalité. Leurs résultats ont montré que ce sont les inférences vis-à-vis de l'intentionnalité qui s'activaient le plus rapidement ($M = 1490$ ms)²¹. De nombreux autres travaux ont mis en avant ce besoin fondamental de l'être humain à inférer rapidement les états mentaux d'autrui (e.g., Gergely et al., 1995 ; Hassin et al., 2005 ; Liu et al., 2017 ; Moskowitz & Olcaysoy Okten, 2016). Ainsi, il semblerait que, loin d'être un raisonnement lent et rationnel, la perception de l'intentionnalité s'active automatiquement à travers des réseaux neuronaux dédiés (Pelphrey et al., 2004).

Dans *La Psychologie des relations interpersonnelles*, Heider (1958) expliquait cette volonté affirmée de détecter les actions humaines volontaires en soulignant son caractère adaptatif : « Lorsque je suis menacé par un danger provenant d'une source impersonnelle, il me suffit généralement de modifier les conditions afin d'échapper au danger. Si je suis menacé par des chutes de pierres sur une montagne, je peux sortir de la zone dangereuse et chercher un

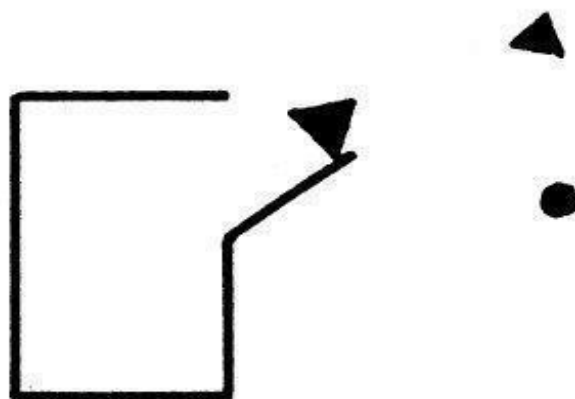
²⁰ Pour la vidéo visionnée par les participants :
https://www.youtube.com/watch?v=VTNmLt7QX8E&ab_channel=Kenjirou

²¹ Pour des résultats similaires en neurosciences, voir Decety et Cacioppo (2012)

abri. Les pierres ne changeront pas de trajectoire pour me retrouver derrière l'abri. Mais si une personne veut me frapper avec une pierre et qu'elle court plus vite que moi, je suis beaucoup plus exposé au danger d'être frappé et je dois utiliser des moyens très différents pour ne pas être frappé : Je peux la frapper en retour et la mettre hors d'état de nuire avant qu'elle ne m'ait frappé, je peux demander grâce, ou je peux essayer de me déplacer de telle sorte qu'elle ne saura pas où je suis passé. » (p.101). Heider développe ici l'idée selon laquelle nous devons savoir discerner rapidement si le dommage en question a été provoqué par un agent humain ayant de mauvaises intentions à notre égard (causalité personnelle) ou non (causalité impersonnelle). En plus de ce caractère adaptatif immédiat, la détection d'intentions malveillantes permet d'effectuer des jugements moraux et de réguler le comportement des individus dans l'espace social. Alors que le raisonnement contrefactuel permettrait d'établir l'*actus reus* (i.e., lien causal entre une action et un dommage), l'attribution d'états mentaux permettrait d'établir le *mens rea* (i.e., les états mentaux de l'agent à l'origine de l'action). Cette deuxième dimension est un objectif tout aussi crucial pour notre survie et notre adaptation que celui de comprendre les liens causaux stricts d'une situation en termes de nécessité réelle et de sensibilité.

Figure 5

Capture d'écran de la vidéo présentée dans l'étude de Heider et Simmel (1944).



Les évaluations négatives spontanées

Selon Alicke (2000), lorsque nous détectons qu'un individu est doté d'intentions ou de motivations malveillantes, ou qu'il a enfreint une norme morale volontairement, nous allons éprouver une évaluation spontanée négative à son encontre. Cette évaluation peut être provoquée par les états mentaux de l'agent (intentions, motivations, etc.), mais également par des facteurs extralégaux²² tels que la réputation de l'agent, son attractivité sociale, son origine ethnique, son genre ou la sévérité des conséquences de ses actions (Alicke, 2000 ; Alicke et al., 1990, 2008 ; Alicke & Zell, 2009). Ces évaluations spontanées négatives comprennent des jugements attitudinaux négatifs sur l'événement et les individus impliqués, ainsi que des réponses émotionnelles qui modulent la force de ces évaluations (Alicke, 2000, p.564). En effet, des émotions négatives s'activent dès lors que nous détectons la présence d'états mentaux malveillants (e.g., Kédia et al., 2008 ; Russell & Giner-Sorolla, 2011 ; Umphress et al., 2013). Toujours selon Alicke (2000), les évaluations spontanées négatives activent le désir de blâmer la ou les personne(s) qui évoque(nt) l'affect le plus négatif. Plus l'évaluation spontanée envers l'agent qui transgresse une norme morale sera négative, plus la motivation de le blâmer sera importante. Ce désir de blâmer aurait pour fonction de publiquement réguler le comportement des membres d'une communauté qui enfreignent des normes morales avec des états mentaux pernicieux (Malle et al., 2014 ; Tetlock, 2002). Le blâme permettrait ainsi d'attirer l'attention de la communauté sur les méfaits d'un agent, et constituerait une forme symbolique de sanctions (Alicke, 1992).

En somme, l'activation d'évaluations spontanées négatives et d'une volonté d'en blâmer la source est endémique aux contextes qui nous intéressent (i.e., situations à l'issue négative impliquant un agent humain), notamment lors d'infraction de normes morales (Voiklis

²² Alicke utilise le terme *facteurs extralégaux* pour désigner tous les facteurs qui, selon le droit, ne doivent pas intervenir dans l'évaluation de la culpabilité de l'accusé (Alicke et al., 2008 ; Alicke & Zell, 2009).

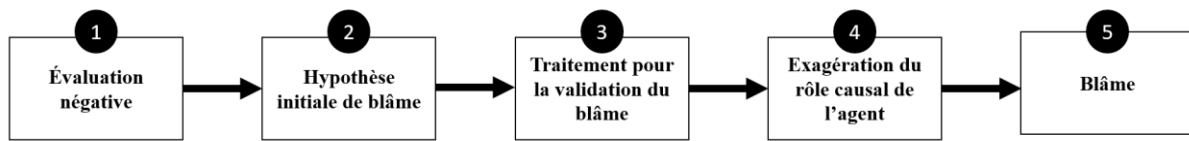
& Malle, 2018). Contrairement aux études sur les boules de billard et les ballons de football des modèles contrefactuels, les situations qui nous intéressent possèdent cette dimension fondamentalement évaluative. Or, de nombreuses données indiquent que ces évaluations et émotions peuvent biaiser le raisonnement (pour une revue, voir Blanchette et al., 2018), et notamment les attributions de responsabilité et de blâme (Burger, 1981 ; Feigenson, 2016 ; Feigenson & Park, 2006). La question qui se pose maintenant est : ces évaluations spontanées négatives peuvent-elles également biaiser le jugement causal ? Nous verrons dans la prochaine partie que la prise en compte de cette dimension sociale, morale et évaluative, et tout ce qu'elle implique en termes de désir de blâmer et de sanctionner, rend crédible l'hypothèse d'un raisonnement motivé dans les tâches qui nous intéressent.

Le Modèle du Contrôle Coupable et le jugement causal

Selon le Modèle du Contrôle Coupable (Alicke, 2000), le jugement causal de sens commun serait intrinsèquement biaisé par un *processus de validation du blâme*. Nous avons vu que lorsqu'un agent suscite une évaluation spontanée négative, le sens commun va éprouver un désir de le blâmer. Néanmoins, comme les individus préfèrent généralement se considérer comme justes et rationnels, ils cherchent à étayer leurs attributions de blâme par des preuves (Alicke et al., 2011, p.7). Nous avons également vu que pour blâmer et sanctionner un agent pour un dommage, un lien causal suffisamment important entre ses actions et le dommage est requis par la majorité des sociétés et des systèmes juridiques. Par conséquent, pour pouvoir assouvir cette volonté de blâmer, nous aurions tendance à interpréter les preuves disponibles de sorte à soutenir l'hypothèse de blâme et la volonté de blâmer. L'exagération du rôle causal d'un acteur vis-à-vis du dommage est le meilleur moyen de parvenir à cet objectif.

Figure 6

Illustration du processus de validation du blâme d'après Alicke et al. (2011).



Le *processus de validation du blâme* (illustré dans la Figure 6, ci-dessus) débute par une évaluation spontanée négative envers l'agent impliqué (1). Cette évaluation conduirait à l'hypothèse selon laquelle la source de l'évaluation est blâmable, et à un désir de blâmer cette source (2). Ce désir amènerait les observateurs à traiter les informations de manière biaisée afin de soutenir cette hypothèse de blâme (3), et ce, par trois moyens différents : assouplissement des critères pour blâmer (i), altération de la perception des preuves (ii) et recherche biaisée d'informations (iii) (Alicke, 2000). Premièrement (i), les individus accepteraient un lien causal plus faible pour pouvoir blâmer l'agent par rapport à ce qui est normalement attendu. Deuxièmement (ii), les individus percevraient un lien causal plus important entre l'agent blâmable et le dommage. Enfin (iii), similaire au biais de confirmation d'hypothèse (Wason, 1960), les individus motivés à blâmer chercheraient activement tous les éléments leur permettant de renforcer le lien causal entre l'agent et le dommage. Ces trois stratégies permettraient d'exagérer le rôle causal de l'agent vis-à-vis du dommage (4). Une fois son rôle causal amplifié, l'individu aurait suffisamment d'éléments pour pouvoir justifier le blâme qu'il souhaite attribuer à l'agent ayant suscité l'évaluation négative initiale (5).

La première étude ayant testé l'hypothèse du *processus de validation du blâme* est celle d'Alicke (1992). Les participants devaient lire un scénario dans lequel un individu nommé John roulait au-delà de la vitesse réglementaire et provoquait un accident qui blessait un autre conducteur. Les motivations de John étaient manipulées : John roulait trop vite pour cacher un

CHAPITRE 2

DE L'EFFET DES NORMES À LA VALIDATION DU BLÂME

cadeau d'anniversaire pour ses parents (motivations socialement désirables) ou un sachet de cocaïne (motivations socialement indésirables) qu'il avait laissé en évidence. Dans les deux cas, John enfreint une norme prescrite (i.e., la limitation de vitesse) et les conditions ne diffèrent pas en termes de critères normatifs de la causalité. En effet, dans les deux conditions, les actions de John sont une condition nécessaire réelle à l'accident (i.e., l'accident n'aurait pas eu lieu sans son excès de vitesse), et ne varient pas en termes de sensibilité car les circonstances de fond sont les mêmes (e.g., il roule à la même vitesse quelles que soient ses motivations). Il était ensuite demandé aux participants de compléter la phrase suivante : « La cause principale de l'accident était _____ ». Les résultats ont montré que John était davantage désigné comme cause de l'accident lorsque ses motivations étaient socialement indésirables (vs. désirables). Les modèles contrefactuels ne permettent pas d'expliquer ce résultat puisque dans les deux cas, le point d'intervention se trouve sur le comportement anormal de John en excès de vitesse. Il semblerait ici que ce soit la valence morale des motivations de John, à savoir la motivation de cacher un sachet de cocaïne plutôt qu'un cadeau, qui explique l'augmentation de la sélection causale de John. Cet effet de la valence morale sur le jugement causal a été corroboré à plusieurs reprises par la suite (Alicke et al., 2008, 2011 ; Lagnado & Channon, 2008 ; Phillips & Shaw, 2015 ; Rogers et al., 2019). Par exemple, Alicke et ses collègues (2011) ont mis en évidence que les participants attribuaient un score de causalité plus élevé à un individu qui tuait une victime socialement attractive (un médecin) plutôt que non attractive (un pédophile sortant de prison). Ces mêmes auteurs ont également répliqué l'étude de la *pen vignette* en demandant aux participants les raisons de la sélection du Professeur comme cause. Sur les 71 participants interrogés, 70 ont désigné le caractère blâmable de son comportement. Aucun n'a désigné l'infraction d'une norme.

Tous ces travaux semblent aller dans la même direction : il semblerait que nous soyons face à un effet plus large que celui des normes. Les évaluations spontanées négatives,

caractéristiques des situations qui nous intéressent dans cette thèse, pourraient être à l'origine, non seulement de l'effet des normes, mais de chaque effet de l'évaluation négative d'un élément de la situation (i.e., agent, victime, conséquences) sur le jugement causal.

Distinction entre raisonnement contrefactuel et CCM

Le CCM s'accorde avec les modèles contrefactuels sur le fait que les violations de normes sont les principaux déterminants du jugement causal pour les événements qui n'impliquent pas d'évaluation vis-à-vis d'agents humains (e.g., une voiture qui refuse de démarrer ou une panne de courant) (Alicke et al., 2011, p.7). En revanche, le désaccord réside dans la non-prise en compte des variables motivationnelles par l'approche contrefactuelle (i.e., la volonté de blâmer) lorsqu'il s'agit de l'infraction d'une norme morale par un agent humain. Les modèles contrefactuels considèrent que l'individu raisonne de manière rationnelle et, tel un « ingénieur » (Alicke et al., 2011, p.29), vise à corriger les événements anormaux de son environnement. Les deux approches s'accordent sur le fait que le sens commun se comporte ainsi lorsqu'un événement enfreint une norme statistique par exemple. Néanmoins, lorsqu'une dimension évaluative est présente en raison de la valence morale de la situation, les deux approches divergent.

L'explication des modèles contrefactuels de l'effet des normes « ne mentionne aucune sorte de jugement moral » (Hitchcock & Knobe, 2009, p.25). Les dimensions évaluatives et affectives sont totalement écartées car le jugement causal y est considéré comme parfaitement rationnel et pragmatique (i.e., en se focalisant sur les interventions efficaces et pertinentes). Alicke (2000) reconnaît que les simulations contrefactuelles (nommées *analyses du contrôle causal effectif* dans son modèle) jouent un rôle dans l'évaluation du lien causal entre l'agent et le dommage. En revanche, il considère que ces simulations n'ont pas besoin de l'infraction d'une norme pour opérer. Celles-ci se font par défaut et sont biaisées lorsque des évaluations

spontanées négatives surviennent. C'est sur ce point que les deux modèles divergent de façon considérable.

Dès à présent, nous privilégierons l'emploi de « l'effet de la valence morale » dans le présent manuscrit plutôt que celui de « l'effet des normes ». Bien que l'utilisation de l'un ou l'autre des termes renvoie à un parti pris théorique, l'utilisation de ce terme est motivée par deux raisons. La première est que le CCM posséderait un pouvoir explicatif plus important que les modèles contrefactuels pour les situations qui nous intéressent (i.e., situations à l'issue négative impliquant un agent humain). En effet, nous avons vu que l'effet des normes ne serait qu'une dimension d'un effet plus large, à savoir celui de la valence morale de l'agent sur le jugement causal. Effectivement, les évaluations négatives en jeu dans le *processus de validation du blâme* peuvent être provoquées par d'autres facteurs que les normes morales, tels que la réputation, l'attractivité sociale, l'ethnie de l'agent, les préjugés ou la nature positive ou négative des conséquences. Néanmoins, il convient de préciser que la supériorité du pouvoir explicatif du CCM ne concerne que les événements impliquant une issue négative et un agent humain, car ce modèle ne propose pas d'explication pour l'effet des infractions des normes statistiques (e.g., court-circuit) vu précédemment.

La deuxième raison à cela est qu'il s'agit du modèle dont les prédictions ont les plus grandes conséquences d'un point de vue sociétal. Comme nous l'avons vu, le CCM considère que l'évaluation des critères normatifs précédemment décrits est intrinsèquement biaisée par les évaluations affectives et morales des individus. Lorsque des évaluations négatives sont présentes, le jugement causal de sens commun serait influencé par un *processus de validation du blâme* (i.e., exagérer le lien causal qui unit un agent et un dommage pour assouvir notre volonté de le blâmer). Or, nous avons vu que les professionnels du droit partent du postulat d'un sens commun rationnel. Des travaux ont pourtant mis en évidence l'effet de la valence

morale sur les jugements non seulement du sens commun, mais également de professionnels du droit (Kneer & Bourgeois-Gironde, 2017 ; Tobia, 2020). En cela, le CCM est probablement le modèle permettant le mieux d'expliquer les erreurs judiciaires liées à l'effet des préjugés ou autres évaluations affectives négatives sur le jugement causal.

Conclusion

Dans ce chapitre, nous avons vu différentes approches théoriques sur les jugements causaux de sens commun dans des situations à l'issue négative impliquant un agent humain. D'un côté, les modèles de « l'Homme comme avocat » et du raisonnement contrefactuel soutiennent que le sens commun est rationnel dans ces situations. Et pour cause, celui-ci respecte les critères normatifs de la causalité : nécessité réelle et sensibilité. Concernant l'infraction d'une norme, celle-ci détermine le point d'intervention le plus pertinent pour contrôler l'environnement. L'individu est ici perçu comme un « ingénieur » rationnel dont le seul objectif est de comprendre et de contrôler son environnement en corrigeant les « anomalies ». A l'opposé, le CCM postule qu'un objectif central pour le sens commun, dans les situations à l'issue négative impliquant un agent humain, est de détecter et de sanctionner les individus blâmables. Cet objectif occuperait une place tellement importante dans la cognition humaine que le jugement causal serait, par défaut, intrinsèquement biaisé par les évaluations spontanées négatives générées par un individu blâmable. Dans le prochain chapitre, nous nous intéresserons à l'articulation de l'effet de la valence morale sur le jugement causal avec les critères normatifs du jugement causal vus dans le Chapitre 1. Ceci constitue alors notre seconde grande question : le *processus de validation du blâme* irait-il jusqu'à la neutralisation du respect du critère de nécessité réelle ? et du critère de sensibilité ?

Chapitre 3. Délimiter l'effet de la valence morale sur le jugement causal

« What is needed, from our perspective, is an analysis that distinguishes different kinds of reasoning errors. »

Reyna, Lloyd et Brainerd (2003, p.218)

De prime abord, la question de savoir si le jugement causal est rationnel ou non semble avoir déjà été tranchée par les travaux présentés dans le chapitre précédent. En effet, l'hypothèse d'une parfaite adéquation entre les critères normatifs et les jugements causaux de sens commun n'apparaît pas soutenable en l'état actuel des connaissances. La prise en compte de facteurs extra-normatifs, comme la valence morale, suffit à qualifier le *processus de validation du blâme* de biais de jugement. Un biais cognitif étant un écart entre le jugement de sens commun et un modèle normatif, l'amplitude de cet écart peut varier d'un biais à l'autre. Il serait alors pertinent de remplacer une conception dichotomique de la rationalité (i.e. un jugement est biaisé ou non) par la notion de rationalité graduée (Brainerd & Reyna, 2001 ; Etzioni, 2014), puisqu'il existerait des biais de natures et de « gravité » différentes (Brainerd & Reyna, 2001 ; Fischhoff, 1982 ; Larrick, 2004 ; Stanovich, 2018). Cette diversité permet d'expliquer notamment pourquoi la capacité à penser de façon rationnelle sur une tâche n'apparaît pas ou peu généralisable à d'autres tâches (Barnett & Ceci, 2002 ; Ellis, 2018 ; Lilienfeld et al., 2009 ; Willingham, 2008). Par conséquent, les méthodes de débiaisement²³ devraient être appliquées « sur mesure » en prenant en compte les spécificités du domaine et du biais étudié. Connaître la nature et les conditions d'apparition du biais étudié permettrait

²³ Par *méthode de débiaisement*, nous entendons toute technique permettant de supprimer ou de réduire l'intensité ou la fréquence des biais (Lilienfeld et al., 2009).

alors de déterminer la méthode de débiaisement la plus appropriée (Arkes, 1991 ; Babai et al., 2015 ; Fischhoff, 1982 ; Larrick, 2004 ; Soll et al., 2015).

Dans le domaine qui nous intéresse, plusieurs hypothèses permettraient d'expliquer l'écart observé entre le jugement causal de sens commun et les critères normatifs. Une première explication pourrait être que les individus ne connaissent pas les critères normatifs du jugement causal ou qu'ils ne les mobilisent pas lors de la tâche. Dans ce cas, la stratégie la plus efficace consisterait en une méthode d'apprentissage permettant aux individus d'intégrer et d'appliquer ces critères (e.g., proposer une formation sur le jugement causal aux jurés). Avec les travaux sur le raisonnement contrefactuel, nous avons cependant vu que les individus semblent capables de déterminer spontanément la présence d'un lien causal (i.e., nécessité réelle) et d'en déterminer sa force (i.e., sensibilité). Ainsi, le problème ne semble pas résider dans l'intégration ou l'application de ces critères normatifs. Selon le CCM, l'irrationalité des jugements causaux reposerait sur une interférence entre deux processus : le *processus de validation du blâme* et l'évaluation objective des liens causaux. En d'autres termes, les individus produiraient un jugement causal rationnel en appliquant les critères normatifs pertinents, lequel serait parasité par un autre processus lorsqu'une réaction évaluative négative entre en jeu. Il reste à déterminer dans quelle mesure ce second processus altère le premier. Autrement dit, quel est l'écart entre les jugements causaux de sens commun et les critères normatifs dans les situations propices au *processus de validation du blâme* ? La réponse à cette question sera l'objectif du Chapitre 3.

Comme nous l'avons vu, l'évaluation des critères normatifs et la validation du blâme ont des objectifs distincts : comprendre et prédire l'environnement pour l'évaluation des critères normatifs, et, blâmer les individus malveillants pour la validation du blâme. Ces deux objectifs peuvent être incompatibles, par exemple lorsqu'un individu malveillant n'est pas une condition nécessaire réelle au dommage. Si le *processus de validation du blâme* submerge

l'évaluation rationnelle des liens causaux, est-il envisageable que les individus aillent jusqu'à inventer un lien causal inexistant entre l'agent malveillant et le dommage ? En revanche, si c'est l'application des critères normatifs qui prime, les individus ne désigneraient pas l'agent malveillant comme cause en l'absence d'un lien causal, et ce, malgré une forte motivation à le blâmer.

En d'autres termes, la question plus générale à laquelle ce chapitre va tâcher de répondre est la suivante : Dans quelle mesure les jugements causaux de sens commun s'écartent-ils des critères normatifs de la causalité dans les situations à l'issue négative impliquant un agent humain ? Cette question peut être scindée en deux aspects distincts : l'écart des jugements de sens commun par rapport (1) au critère de nécessité réelle et (2) par rapport au critère de sensibilité. Répondre à ces questions permettrait de déterminer la place que le jugement causal de sens commun occupe sur le continuum de la rationalité et de mieux cerner les méthodes de débiaisement à mettre en place pour en limiter les impacts sociétaux²⁴. Nous allons désormais voir que la littérature ne permet pas de répondre à ces questions, que ce soit sur le critère de nécessité réelle ou sur le celui de sensibilité. La série d'études que nous avons menées et que nous présenterons ensuite, quant à elle, ambitionnait d'apporter de premiers éléments de réponse.

Processus de validation du blâme et nécessité réelle

Si le *processus de validation du blâme* pouvait aller jusqu'à la transgression du critère de nécessité réelle, ceci dénoterait une irrationalité majeure de la part des individus. De plus, cela se traduirait par des conséquences sociétales considérables (e.g., désigner un innocent coupable d'un dommage qu'il n'a pas commis). Les prédictions du CCM ne sont pas toujours

²⁴ Il convient de noter qu'aucune méthode de débiaisement n'a été testée à ce jour pour réduire l'effet de la valence morale sur le jugement causal.

consistantes sur cette question. D'un côté, Alicke et ses collègues (2011) indiquent que la validation du blâme consiste à exagérer le contrôle réel ou « potentiel » d'un agent sur un dommage (p.675). Or, désigner un agent comme cause alors que sa contribution n'est que potentielle serait une transgression du critère de nécessité réelle. D'un autre côté, ces mêmes auteurs indiquent que les individus n'iraient pas jusqu'à inventer un lien causal inexistant : « les effets des réactions évaluatives seront annulés lorsque le rôle causal d'un agent est sans ambiguïté, c'est-à-dire lorsqu'un agent est clairement, ou clairement pas, un contributeur important à l'événement » (Alicke, 2011, p.685). A notre connaissance, il existe trois études ayant fait varier la valence morale et la présence d'un lien causal (Alicke, 1992, Etude 4 ; Alicke et al., 2011, Etude 3 ; Hilton et al., 2021, Etude 3). Néanmoins, nous allons voir qu'aucune ne permet de savoir si les participants iraient jusqu'à inventer un lien causal pour blâmer l'agent impliqué.

La première étude ayant fait varier à la fois la nécessité réelle des actions de l'agent et sa valence morale est celle d'Alicke (1992, Étude 4). Le scénario présentait un homme qui empoisonnait son associé (intentionnellement vs. accidentellement). Dans une condition, l'associé mourrait directement du poison (i.e., condition nécessaire), et dans une autre condition, le médecin affirmait que l'associé était mort subitement d'une maladie du foie et que le poison n'était pas suffisamment concentré pour le tuer (i.e., condition non nécessaire). Néanmoins, seuls les effets principaux de ces deux variables ont été testés sur les jugements causaux. L'effet d'interaction n'étant pas étudié²⁵, les résultats ne nous permettent pas de déterminer si l'effet de la valence morale se manifestait dans la condition où l'agent n'était pas une condition nécessaire réelle au dommage. Plus tardivement, Alicke et ses collègues (2011, Étude 3) ont cherché à tester directement cette interaction.

²⁵ Les données ne sont pas accessibles pour nous permettre de faire une réanalyse.

Ils ont utilisé un scénario dans lequel un agent tirait et tuait une victime qui s'était introduite chez lui. Les auteurs ont d'abord fait varier la caractérisation de la victime, présentée comme un ancien détenu dangereux qui s'était introduit dans la maison de l'agent (caractérisation négative), ou un médecin qui était entré dans la maison pour nourrir le chat à la demande de la femme de l'agent alors qu'elle était absente (caractérisation positive). Puis ; ils ont manipulé le modèle causal à travers trois conditions : l'agent tirait sur la victime et la tuait instantanément (condition « tir »); l'agent tirait sur la victime et la tuait instantanément mais celle-ci serait morte d'une tumeur cérébrale au bout de deux semaines (condition « tir + tumeur cérébrale ») ; l'agent tirait sur la victime et la tuait instantanément mais celle-ci a souffert d'un anévrisme cérébral presque au même moment où elle a été touchée par la balle (condition « tir + anévrisme »). Dans les deux premières conditions, le tir de l'agent était la seule condition nécessaire à la situation immédiate. Dans la troisième, la mort de la victime (i.e., dommage) aurait été causée par l'anévrisme presque immédiatement après le tir si celui-ci n'avait pas eu lieu. Alicke et ses collègues ont émis l'hypothèse que les effets des réactions évaluatives négatives à l'égard de l'agent qui tire sur une victime, caractérisée positivement vs. négativement, seraient annulés lorsque le rôle de l'agent n'est pas clairement lié au dommage (c'est-à-dire la condition « tir + anévrisme »). Les résultats ont confirmé leur hypothèse. Sur la base de ces résultats, les évaluations spontanées négatives peuvent donc influencer le jugement causal, dans des circonstances respectant toutefois le critère de nécessité.

Malgré les contributions scientifiques de cette étude, l'opérationnalisation de la condition « tir + anévrisme » par rapport au critère de nécessité réelle peut néanmoins encore être débattue. D'une part, c'est bel et bien l'agent qui a tué la victime dans la situation réelle. En effet, avant de parler de l'anévrisme qui aurait tué la victime sans l'action de l'agent, le scénario précise « « la balle a atteint Poole [la victime] à la poitrine et au dos et l'a tué presque instantanément » (p.248). Par conséquent, sa mort n'aurait pas eu lieu telle qu'elle a réellement

eu lieu si l'agent n'avait pas tiré. Contrairement à leurs attentes, les auteurs semblent donc avoir utilisé un modèle causal disjonctif (e.g., Icard et al., 2017 ; Kominsky et al., 2015 ; Kominsky & Phillips, 2019), c'est-à-dire une situation dans laquelle deux actions sont suffisantes pour provoquer le dommage, plutôt qu'un modèle dans lequel il n'y a pas de lien causal entre l'agent et le dommage. D'autre part, dans cette condition, les participants ont attribué un score moyen de causalité de 5.10 lorsque la victime est caractérisée positivement et de 4.80 lorsque celle-ci est caractérisée négativement. L'échelle utilisée dans cette étude étant en 9 points (1 = pas du tout la cause ; 9 = totalement la cause), ces résultats ne permettent pas de conclure qu'une majorité de participants considérait qu'il n'y avait pas de lien causal entre l'agent et le dommage. Afin de réellement étudier l'articulation des effets de la valence morale d'un agent avec le critère de nécessité réelle, il convient de tester un modèle causal dans lequel il n'y a réellement pas de lien causal entre l'agent et le dommage.

Enfin, dans une étude que nous avons menée (Hilton et al., 2021, Étude 3), les participants devaient lire le scénario suivant : « Suzy et Billy ont chacun un ballon et jouent dans un parc. Billy décide de viser une peluche placée au bord de l'étang alors que Suzy décide de viser un arbre. Le ballon de Suzy est dévié et vient en fait toucher en premier la peluche qui tombe dans l'étang. Billy avait par ailleurs parfaitement tapé dans son ballon qui aurait lui aussi touché la peluche et l'aurait ainsi fait tomber dans l'étang si Suzy ne l'avait pas fait. ». Il s'agit ici d'un modèle causal dans lequel l'agent qui a l'intention du dommage (i.e., Billy) n'est pas une condition nécessaire réelle au dommage. Les résultats ont montré qu'aucun des 42 participants n'ont désigné Billy comme cause de la chute de la peluche²⁶. Malgré la présence d'un agent qui a cherché à provoquer le dommage intentionnellement, ces résultats vont dans le sens d'un respect du critère de nécessité réelle de la part des participants. La principale

²⁶ 2 participants ont malgré tout indiqué que les deux agents (Billy + Suzy) étaient la cause de la chute de la peluche.

critique que l'on peut apporter à cette étude est qu'il s'agit d'une situation dont la gravité est relativement faible. Selon le CCM, l'évaluation spontanée négative doit être suffisamment forte pour que le *processus de validation du blâme* soit engagé (Alicke et al., 2011, p.675). Par conséquent, il apparaît nécessaire d'utiliser un scénario dont les événements sont suffisamment graves pour provoquer un minimum d'évaluation spontanée négative.

En somme, aucune de ces trois études ne semble apporter de preuves suffisamment convaincantes à la question de l'articulation entre l'effet de la valence morale et le critère de nécessité réelle. Pour y remédier, plusieurs conditions sont requises. Premièrement, l'étude devra s'intéresser précisément à l'interaction entre la valence morale et la nécessité réelle. Deuxièmement, le critère de nécessité réelle devra être manipulé sans ambiguïté : la négation de l'action de l'agent ne doit pas empêcher le dommage tel qu'il s'est réellement produit. Enfin, la situation doit être suffisamment grave pour susciter une évaluation spontanée négative suffisamment forte. Pour ce faire, il conviendrait par exemple d'utiliser une situation relevant d'une juridiction pénale (e.g., crime). Pour chaque étude présentée dans ce chapitre, nous avons veillé au respect de ces trois conditions.

Processus de validation du blâme et sensibilité

La seconde question qui se pose est celle de l'articulation entre le critère de sensibilité (i.e., force causale) et l'effet de la valence morale. Contrairement au critère de nécessité réelle, l'articulation entre la sensibilité et la valence morale d'un agent n'a, à ce jour, pas fait l'objet d'investigation. Il serait donc nécessaire d'y remédier. Une étude qui ambitionnerait de répondre à cette question se devrait néanmoins de porter une attention particulière à une ambiguïté théorique et méthodologique, à savoir la confusion possible entre l'état épistémique de l'agent (e.g., son intention de provoquer le dommage) et sa sensibilité.

Nous avons vu précédemment que l'intentionnalité de l'agent est un facteur clé pour faire varier sa valence morale. Selon le CCM, si un agent provoque un dommage intentionnellement (vs. non intentionnellement), il sera davantage considéré comme cause de celui-ci en raison de réactions évaluatives négatives à son égard. Néanmoins, une autre hypothèse en lien avec la sensibilité a été avancée pour expliquer cet effet (Lombrozo, 2010 ; Murray & Lombrozo, 2017). Celle-ci précise qu'un agent qui a l'intention de provoquer un dommage est moins sensible aux circonstances de fond de l'événement, qu'un individu ne le souhaitant pas (i.e., en le provoquant de manière accidentelle par exemple). En d'autres termes, l'intentionnalité n'agirait pas uniquement sur la valence morale de l'agent, mais également sur sa force causale. Ceci est notamment lié au fait qu'un agent qui agit intentionnellement va s'adapter aux variations des circonstances pour accomplir son objectif. Comme dans l'exemple de Heider (1958) que nous avons déjà évoqué (voir pp. 56-57), il est plus facile de se protéger d'un éboulement en modifiant les circonstances de fond (e.g., s'écarter de la zone dangereuse) que d'un individu déterminé à nous frapper. Nous retrouvons un autre exemple de ceci dans l'étude 4 d'Alicke (1992). Ici l'intentionnalité de l'agent permettait d'opérationnaliser sa valence morale avec : dans le scénario « intentionnel », l'agent empoisonnait volontairement la victime ; dans le scénario « non intentionnel », l'agent avait bu trop d'alcool et empoisonna la victime en confondant involontairement un colorant toxique de sa femme chimiste avec un bouillon. Dans ce second scénario, plusieurs variations des circonstances annuleraient le lien causal entre l'agent et le décès de la victime, notamment la quantité d'alcool ingéré par l'agent ou la présence du colorant toxique dans le réfrigérateur. En revanche, dans le scénario « intentionnel », peu de variations des circonstances de fond annuleraient le lien causal de l'agent déterminé à provoquer le décès de la victime. L'hypothèse d'une différence de force causale entre les deux conditions est donc défendable dans ce type de situation. Cette ambiguïté entre intentionnalité et force causale peut ainsi engendrer des conclusions erronées (e.g.,

conclure à un effet de la valence morale alors que c'est la force causale qui a influencé le jugement causal).

Une solution possible pour résoudre ce problème serait de maintenir l'intentionnalité de l'agent constante et d'opérationnaliser la valence morale de l'agent en faisant varier ses motivations (i.e., la raison de son intention à produire son action) par exemple. En maintenant constante l'intentionnalité vis-à-vis du dommage et en variant les motivations de l'agent, nous nous assurons que l'effet observé sur le jugement causal peut être attribué à un effet de la valence morale, plutôt qu'à la prise en compte du critère normatif de force causale. Par ailleurs, les motivations de l'agent constituent un facteur déterminant pour les réactions évaluatives des observateurs, car elles permettent d'indiquer les raisons (moralement bonnes ou mauvaises) d'une intention ou d'une action (Carlson et al., 2022 ; Malle et al., 2014). Ainsi, dans nos études, nous avons fait varier ce facteur avec des motivations « louables » versus « blâmables », afin de provoquer une évaluation spontanée en termes de « bien-mal » chez les participants. Cette précaution nous permettra de distinguer sans ambiguïté ce qui relève d'un effet de la valence morale sur le jugement causal (i.e, biais de jugement), de ce qui relève de l'effet de la sensibilité sur le jugement causal (i.e., analyse rationnelle de la force causale).

Études 1 et 2 : Effet de la valence morale sur le jugement causal en fonction des critères normatifs de nécessité réelle et de sensibilité

Ces deux études ont été acceptées pour une publication dans la revue *Philosophical Psychology*, dont la référence est précisée ci-après. Des analyses statistiques bayésiennes viennent ici compléter certaines analyses présentées dans l'article.

Goulette, V., & Verkampt, F. (forthcoming). Blame-validation: Beyond rationality? Effect of causal link on the relationship between evaluation and causal judgment. *Philosophical Psychology*. <https://doi.org/10.1080/09515089.2023.2217834>

Étude 1 : Effet de la valence morale sur le jugement causal en fonction des critères normatifs de nécessité réelle et de sensibilité avec une question ouverte

L'objectif principal de cette étude était d'examiner l'articulation entre l'effet de la valence morale sur le jugement causal, en lien avec les critères normatifs de nécessité réelle et de sensibilité. D'une part, il s'agissait de déterminer si la volonté de blâmer un agent malveillant pouvait inciter les individus à inventer un lien causal inexistant entre cet agent et le dommage en question. Si tel est le cas, ceci constituerait une violation d'un critère fondamental, pour les approches normatives en philosophie et en droit, à savoir le critère de nécessité réelle. D'un point de vue appliqué, ceci signifierait notamment que les jurés pourraient désigner un accusé comme causalement responsable d'un dommage que celui-ci n'a pas commis. D'autre part, il s'agissait d'étudier l'impact de la force causale du lien unissant l'agent au dommage sur l'apparition de l'effet de la valence morale sur le jugement causal. Ceci permettra d'en apprendre davantage sur les conditions d'apparition de cet effet, et de tester l'hypothèse du CCM selon laquelle l'ambiguïté d'un lien causal est favorable à son apparition. Pour ce faire, les participants de notre étude étaient invités à lire un scénario nous permettant d'opérationnaliser la valence morale de l'agent (motivations de l'agent : louables, blâmables),

ainsi que le modèle causal du scénario (lien causal entre l'agent et le dommage : fort, faible ou préempté).

Notre première hypothèse était que les individus sélectionneraient la cause réelle d'un dommage en se basant sur les deux critères normatifs du jugement causal. Plus précisément, l'agent devrait être sélectionné comme cause du dommage si son action était une condition nécessaire réelle à l'émergence de ce dommage (*lien fort* et *lien faible* vs. *lien préempté*). De plus, l'agent devrait être davantage sélectionné comme cause lorsque son action était insensible (vs. sensible) aux circonstances de fond (*lien fort* vs. *lien faible*). Cependant, selon le CCM, les motivations blâmables de l'agent devraient entraîner des évaluations spontanées négatives qui pourraient avoir un impact sur la rationalité des jugements causaux. Par conséquent, en considérant l'interaction entre le lien causal et les motivations de l'agent, nous prédisions que l'effet de la valence morale ne devrait se produire que lorsque le rôle causal de l'agent était ambigu (*lien faible* vs. *lien fort* et *lien préempté*). En effet, selon le CCM, l'effet de la valence morale devrait être neutralisé lorsque l'agent est clairement (i.e., *lien fort*) ou n'est clairement pas (i.e., *lien préempté*) un contributeur important du dommage (Alicke, 2011, p.685). Ceci implique que si le lien causal entre l'agent et le dommage est trop fort, l'agent sera désigné comme cause peu importe ses motivations. A contrario, si ce lien est inexistant, l'agent ne sera pas désigné comme cause peu importe ses motivations.

Méthode

Design. L'expérience était un plan 3 (Lien : fort, faible, préempté) × 2 (Motivations : louables, blâmables) en inter-sujets.

Participants. Pour estimer la taille d'échantillon requise, nous avons conduit une analyse de puissance statistique a priori²⁷. Cette analyse s'est appuyée sur les résultats de la seule étude de l'auteur du CCM ayant utilisé une question ouverte (Alicke, 1992, Etude 1). Les résultats ont indiqué que les données de 39 participants par condition expérimentale permettraient d'obtenir une puissance de 80% avec $\alpha = .05$.

Les participants étaient des étudiants présents sur le campus de l'Université Toulouse Jean Jaurès²⁸. Notre échantillon final comprenait 445 participants (249 femmes, 56%), dont l'âge moyen était de 21.5 ans ($SD = 4.58$; min = 18 ; max = 67). Notre ensemble de données reflétait ainsi une distribution moyenne de 74 participants dans chacune des 6 conditions expérimentales susmentionnées. Tous les participants étaient des volontaires et n'avaient reçu aucune compensation financière pour leur participation. Cette étude a été approuvée par le Comité d'Éthique de la Recherche de Toulouse.

Matériel et procédure. En premier lieu, les participants étaient invités à lire le formulaire de consentement et à indiquer leur accord s'ils l'approuvaient. Une fois leur accord fourni, les participants renseignaient des informations sociodémographiques (genre et âge), puis il leur était demandé de lire attentivement l'un des 6 scénarios distribués aléatoirement. Immédiatement après la lecture du scénario, ils étaient invités à répondre à deux questions

²⁷ L'analyse de puissance a priori a été réalisée avec le package « *pwr* » (version 1.3-0) du logiciel R studio.

²⁸ Une méta-analyse n'a montré aucun effet du type d'échantillon (participants étudiants vs. non étudiants) dans les tâches de « jurés fictifs » pour des affaires d'ordre pénal (Bornstein et al., 2017).

ouvertes (cf. pp.78-79). Un exemple de questionnaire fourni aux participants est disponible en Annexe A.

Scénarios. Les six scénarios décrivaient une femme, Barbara, qui avait l'intention de tuer son mari pour des motivations louables ou blâmables. Les conditions expérimentales sont décrites en détail ci-dessous :

[Motivations louables] « Barbara est sévèrement et continuellement battue par son mari Jean,

[Motivations blâmables] « Barbara est une personne particulièrement vénale, et afin de mettre la main sur l'argent de son mari Jean,

Elle veut le tuer. Au cours d'un repas au restaurant, elle envisage de verser du poison dans le plat de Jean à son insu. A peine après avoir commencé son plat, Jean s'absente un instant pour se rendre aux toilettes.

[Lien causal fort] Durant son absence, Barbara verse le poison dans le plat de Jean, un poison aux effets très rapides. A son retour, Jean poursuit son repas, et décède dans les minutes qui suivent les premières bouchées. »

[Lien causal faible] Durant son absence, Barbara verse le poison dans le plat de Jean. Mais Barbara ne sait pas que le poison qu'elle a mis n'est pas assez fort et ne pourra en aucun cas tuer son mari. A son retour, Jean poursuit son repas, mais le poison provoque un goût tellement désagréable dans son plat qu'après quelques bouchées, Jean décide de commander un nouveau plat. Il s'avère que Jean est extrêmement allergique aux ingrédients contenus dans ce nouveau plat, au point d'en décéder dans les minutes qui suivent les premières bouchées. »

[Lien causal préempté] Durant son absence, Barbara verse le poison dans le plat de Jean. A son retour, Jean est sur le point de poursuivre son repas, lorsqu'une araignée tombe du plafond dans son assiette. Sans avoir repris une seule bouchée de son plat, Jean décide de commander un nouveau plat. Il s'avère que Jean fut extrêmement allergique aux ingrédients contenus dans ce nouveau plat, au point d'en décéder dans les minutes qui suivirent les premières bouchées. »

Pour mesurer les évaluations morales induites par la manipulation des motivations de l'agent, nous avons demandé aux participants d'indiquer, par le biais d'une question ouverte, « Envers qui ou quoi faites-vous le plus de reproches ? ». Les analyses préliminaires soutiennent la pertinence de cette opérationnalisation : les participants mentionnaient davantage Barbara lorsque ses motivations étaient blâmables (61%) plutôt que louables (35%), OR = 2.93, 95% CI [1.99, 4.33], $p < .001$.

Concernant les critères normatifs de nécessité réelle et de sensibilité, ceux-ci ont été opérationnalisés sous forme de lien causal entre l'agent et le dommage structurant les scénarios montrés aux participants. Dans chaque lien causal, l'état épistémique de l'agent était maintenu constant : l'agent avait l'intention de provoquer le dommage, il produisit une action de manière intentionnelle dans le but de provoquer ce dommage, il pensait que le dommage allait survenir, le dommage souhaité est survenu.

Afin d'étudier spécifiquement le respect du critère de *nécessité réelle* dans le jugement causal, nous avons utilisé un lien causal préempté dans lequel l'action de l'agent ne respectait pas le critère de nécessité réelle. Dans ce lien causal, l'agent a produit une action qui aurait provoqué le dommage si une autre condition indépendante de l'agent (i.e., l'araignée qui provoque le changement de plat) ne l'avait pas préempté et causé le dommage. Dans le lien

préempté, il n'y avait donc pas de lien causal réel entre l'action réalisée par l'agent et le dommage souhaité par celui-ci.

Concernant le critère de *sensibilité* (force causale), nous avons décidé de rajouter un élément entre l'action de l'agent et le dommage (i.e., l'allergie aux ingrédients du nouveau plat) dans la condition *lien faible*, afin de rendre la relation entre l'action et le résultat plus sensible aux circonstances de fond. Ceci implique que l'action de l'agent n'aurait pas provoqué le dommage sans cet élément intermédiaire.

Tâche de jugement causal. La tâche de jugement causal était basée sur celle utilisée par Alicke (1992), c'est-à-dire une question ouverte par laquelle les participants étaient invités à identifier la cause principale du dommage (i.e., « La cause principale de cet accident était... »). Conformément à Hilton et ses collègues (2021), cette tâche a été adaptée pour déterminer la cause réelle. Par conséquent, après avoir lu l'un des six scénarios possibles décrits précédemment, les participants devaient répondre à la question suivante : « Quelle est la cause réelle de la mort de Jean ? ».

Codage et notation des données. Les réponses obtenues ont permis de constituer deux variables : « Sélection causale » et « Sélection causale de l'agent ». Pour la variable « Sélection causale », les réponses des participants ont été codées en 5 catégories : Les réponses mentionnant l'agent (e.g., « Barbara » ; 1), son mari (e.g., « Jean » ; 2), une condition physique (e.g., « Poison », « Allergie » ; 3), l'agent et une autre condition (e.g., « Allergie et Barbara », « Jean et Barbara » ; 4), et toute autre condition (e.g., « Araignée », « Restaurant » ; 5). Si les participants se répétaient, la même réponse répétée n'était codée qu'une seule fois. Pour évaluer la fiabilité intercodeurs, deux évaluateurs indépendants (assistants de recherche) ont noté 80 réponses sélectionnées au hasard (soit 18% de l'échantillon). L'accord entre les deux codeurs était de 93,8 % avec un kappa de Cohen de 0,76 (accord substantiel, Landis & Koch, 1977).

Pour la variable « Sélection causale de l'agent », conformément à la procédure d'Alicke (1992), pour évaluer si l'agent était plus sélectionné que toute autre condition, un point a été attribué lorsque la réponse des participants faisait référence à Barbara. Par exemple, lorsqu'ils répondaient « Barbara », « Allergie et Barbara », « Poison et Barbara », le code était « 1 – agent ». Lorsqu'ils ont répondu « Jean », « Poison », « Allergie », « Araignée », aucun point n'était attribué.

Résultats

Condition la plus sélectionnée comme cause réelle (i.e., Sélection causale). En premier lieu, nous avons souhaité savoir quelle cause était majoritairement désignée par les participants dans chaque lien causal. Pour ce faire, nous avons conduit des tests du Khi-deux d'ajustement sur les réponses fournies par les participants pour chaque lien causal (α corrigé à .017). Comme indiqué dans le Tableau 1 (p.81), la distribution des causes réelles rapportées n'était pas uniforme, et ce, quel que soit le lien causal. Nous avons ensuite conduit des tests post-hoc avec correction de Bonferroni pour déterminer si la condition majoritairement mentionnée différait significativement de la seconde condition la plus avancée, et ce, pour chaque lien causal. Les résultats ont montré que, lorsque le lien causal était fort entre l'agent et le dommage, le poison était majoritairement rappelé comme cause réelle par rapport au second facteur le plus sélectionné, à savoir Barbara ($p < .001$). Concernant les liens causaux *lien faible et lien préempté*, c'est l'allergie qui était majoritairement désignée, ($p < .001$) et ($p < .001$) respectivement. Les tailles d'effet étaient toutes très grandes ($w > .50$, Cohen, 1988), ce qui indique un fort consensus entre nos participants.

Tableau 1

Proportions de sélection des différents facteurs comme cause réelle en fonction du lien causal entre l'agent et le dommage

Liens causaux	Type de cause réelle					χ^2 (df)	p	Cohen's w
	Poison /Allergie	Jean	Barbara	Barbara + autre condition	Autre			
Lien fort (n = 150)	.807	.053	.093	.020	.027	347.5 3 (4)	<.00 1	1.52
Lien faible (n = 155)	.755	.058	.026	.013	.148	306.9 (4)	<.00 1	1.41
Lien préempté (n = 140)	.886	.029	.007	.000	.079	414.0 7 (4)	<.00 1	1.72

Sélection de l'agent comme cause réelle (i.e., Sélection causale de l'agent). Dans un second temps, l'occurrence de la réponse « Barbara » comme cause de la mort de son mari a été rentrée comme variable dépendante dans une régression logistique binomiale, avec le lien causal et les motivations comme prédicteurs, ainsi qu'une interaction Motivations \times Lien causal. Ce modèle expliquait 23% de la variance du jugement causal et prédisait les données de manière statistiquement significative, $\chi^2(5) = 34.75, p < .001$.

Les résultats ont montré que le lien causal était un prédicteur statistiquement significatif du jugement causal, $\chi^2(2) = 11.00, p = .004$. Plus précisément, les participants ont plus souvent mentionné l'agent comme cause réelle du dommage lorsque le lien causal entre ses actions et le dommage était fort (11.3 %), plutôt que faible (3.9 %) (OR = 0.30, 95% CI [0.10, 0.76] ; $p = .015$) ou préempté (0.7 %), OR = 0.06, 95% CI [0.01, 0.29] ; $p = .006$. Les motivations étaient également un prédicteur statistiquement significatif du jugement causal, $\chi^2(2) = 9.94, p = .002$. Cependant, l'analyse statistique n'a montré aucune interaction significative entre le lien causal et les motivations sur le jugement causal, $\chi^2(1) = 0.171, p = .918$. Ce résultat non significatif pouvait être lié à la difficulté des modèles de régression logistique à s'adapter aux modalités

qui ne comprennent que des zéros (voir King & Zeng, 2001). Ici, il s'agissait de la condition « lien préempté – motivations louables » (voir Tableau 2, ci-dessous). Les statistiques bayésiennes seraient plus adaptées dans ce genre de cas (Straub et al., 2016)²⁹. Pour mieux rendre compte de ces résultats, nous avons ainsi effectué des régressions logistiques binomiales (α corrigés à .017), ainsi que des analyses bayésiennes pour tables de contingences³⁰, afin de tester l'effet des motivations dans chaque lien causal (voir Tableau 2).

Tableau 2

Proportions de sélections de l'agent comme cause réelle en fonction de ses motivations et du lien causal

Liens causaux	Motivations		OR	<i>p</i>	BF ₁₀
	Louables	Blâmables			
Lien fort (<i>n</i> = 150)	.027	.197	8.85 [2.37, 57.59]	.005	36.81
Lien faible (<i>n</i> = 155)	.013	.063	5.07 [0.79, 98.33]	.143	0.26
Lien préempté (<i>n</i> = 140)	.000	.016	NA	.997	0.06

Note. OR = Odds ratio. Les intervalles de confiance sont entre crochets.

L'effet des motivations sur la sélection causale de l'agent n'était statistiquement significatif que parmi les participants alloués à la condition du lien fort. Comme l'indique le tableau 2, la probabilité qu'un participant sélectionne Barbara comme la cause réelle du dommage dans la condition de lien fort était environ 9 fois plus élevée lorsque ses motivations étaient blâmables vs. louables (grande taille d'effet, Chen et al., 2010). L'intervalle de

²⁹ Par ailleurs, comme évoqué en Introduction, les statistiques bayésiennes permettent de tester l'hypothèse nulle contrairement aux statistiques fréquentistes. Or, nous soutenons l'hypothèse nulle d'absence d'effet de la valence morale dans le lien causal préempté.

³⁰ Tous les facteurs bayésiens rapportés dans cette thèse seront analysés à partir du tableau d'interprétation de Jeffrey (Jarosz & Wiley, 2014 ; Quintana & Williams, 2018). Un facteur bayésien proche de 1 signifie que les données ne permettent pas de départage H0 de H1. Un facteur entre 0 et 1 apporte un soutien à l'hypothèse nulle d'absence d'effet et un facteur supérieur à 1 apporte un soutien à l'hypothèse alternative.

confiance de l'odds ratio excluant le 1, la valeur p très basse et le facteur bayésien indiquant que les résultats sont 36.81 fois plus probables sous l'hypothèse alternative que sous l'hypothèse nulle (soutien fort, $BF_{10} > 30$), nous concluons à un fort effet des motivations sur le jugement causal dans cette condition. Cependant, aucun effet significatif des motivations sur la sélection de l'agent n'a été observé parmi les participants affectés aux deux autres conditions (*lien faible* et *lien préempté*). De façon similaire, les facteurs bayésiens indiquaient un soutien pour l'hypothèse nulle (i.e., absence d'effet des motivations). Ce soutien était modéré ($BF_{10} < .33$) dans la condition du lien faible et fort ($BF_{10} < .1$) dans la condition du lien préempté.

Discussion de l'Étude 1

Les résultats de cette première étude permettent de discerner deux constats majeurs vis-à-vis de la rationalité des jugements causaux de sens commun. Premièrement, les participants ont très majoritairement sélectionné la condition la plus déterminante pour provoquer le dommage (i.e., la moins sensible aux circonstances), à savoir le poison (lien fort) et l'allergie (liens faible et préempté). Il ne s'agissait pas d'une condition pouvant susciter une évaluation négative. Ces résultats entrent en contradiction avec les travaux de Hilton et ses collègues (2010), qui ont corroboré l'hypothèse de Hart Honoré (1985) selon laquelle les individus remontent dans la chaîne causale pour sélectionner une action humaine volontaire comme cause plutôt que des conditions purement physiques (e.g., plaque de glace, tempête). Néanmoins, ces auteurs ont mis en évidence que la suffisance conditionnelle de l'agent (i.e., la probabilité que l'agent provoque le dommage en prenant en compte les éléments précédents de la chaîne) permettait d'expliquer cet effet. La suffisance conditionnelle étant une mesure de force causale³¹, cette préférence pour une action humaine volontaire traduirait plus le respect

³¹ La notion de suffisance conditionnelle a été proposée par Spellman (1997). Selon cet auteur, une cause est une condition qui augmente la probabilité d'apparition d'un résultat.

d'un critère normatif, qu'un effet d'une réaction évaluative négative. Cette interprétation va dans le sens de nos résultats : il semblerait que le critère de force causale soit le plus déterminant pour expliquer la grande majorité des causes réelles sélectionnées par les participants, et ce, alors même qu'une action humaine volontaire et blâmable est présente dans la chaîne causale. Ceci apporte un premier soutien considérable à la défense de la rationalité des jugements causaux de sens commun.

Deuxièmement, conformément aux prédictions du CCM, les critères normatifs ne sont pas les seuls facteurs pris en compte par les participants lorsqu'ils doivent désigner la cause réelle d'un dommage. En effet, bien que minoritaires, certains participants ont sélectionné Barbara, particulièrement lorsque ses motivations étaient blâmables. Cependant, la seconde observation la plus importante qui ressort de nos données est que l'effet de la valence morale n'a été reproduit que dans un seul des trois liens de causalité manipulés, à savoir : celui où non seulement un lien pouvait être établi entre l'action de l'agent et le dommage, mais aussi où ce lien était particulièrement fort et robuste (c'est-à-dire insensible à tout changement des circonstances de fond). Lorsque le lien de causalité entre l'agent humain et le dommage est affaibli ou absent, l'effet de la valence morale sur la sélection causale de l'agent n'apparaît plus. L'étude apporte ainsi deux éclairages principaux sur la relation entre l'effet de la valence morale et les critères normatifs du jugement causal.

D'abord, l'effet des réactions évaluatives négatives sur les jugements causaux (i.e., *processus de validation du blâme*) n'apparaîtrait que lorsque les critères normatifs sont respectés (i.e., lien causal fort). Par conséquent, les résultats soutiennent qu'un agent peut être jugé comme la cause réelle d'un dommage uniquement si ce dernier ne se serait pas produit sans les actions de l'agent, et ce, même si ses motivations étaient blâmables. Ils soutiennent également les discussions sur l'importance de la force causale (i.e., la sensibilité) dans la

construction des jugements causaux (Grinfeld et al., 2020 ; Kominsky et al., 2015 ; Vasilyeva et al., 2018 ; Woodward, 2006). Ce faisant, ces résultats diffèrent légèrement des prédictions basées sur le CCM, qui soutiennent que l'effet de la valence morale serait davantage observé dans des situations ambiguës (Alicke et al., 2011). En effet, le *lien faible* étant plus ambigu que le *lien fort* en ce qui concerne la contribution de l'agent, l'effet de la valence morale aurait dû se manifester principalement dans cette condition.

Ensuite, la seule présence d'un lien causal entre les actions d'un agent blâmable et le dommage (*lien faible vs lien préempté*) n'est pas suffisante pour juger l'agent blâmable comme la cause réelle de cet incident. En ne parvenant pas à reproduire l'effet de la valence morale dans une situation où le lien entre l'agent blâmable et le résultat était faible, nos résultats soutiennent l'hypothèse selon laquelle les individus se préoccupent prioritairement du critère normatif de force causale avant de laisser s'exprimer leur volonté de blâmer.

En résumé, la majorité des participants semblait parfaitement rationnelle en indiquant la condition ayant la plus grande force causale vis-à-vis du dommage. Pour ce qui est de la minorité de participants qui sélectionnaient l'agent humain comme cause, celle-ci s'écartait très peu de cette rationalité en respectant les critères de nécessité réelle et de sensibilité. Ces résultats apportent ainsi un soutien majeur pour la rationalité des jugements causaux, et laissent notamment présager que des jurés n'iraient probablement pas jusqu'à inventer un lien causal inexistant pour blâmer un accusé.

Cependant, l'absence de réplcation de l'effet de la valence morale en présence de liens faible et préempté pourrait être spécifique à la tâche de jugement causal. Comme dans l'étude d'Alicke (1992), les participants de l'étude 1 ont jugé la cause réelle du dommage à l'aide d'une question ouverte, alors que les participants l'ont fait à l'aide de questions à échelle dans la plupart des recherches qui ont reproduit l'effet de la valence morale (e.g., Alicke en 1992,

Études 3 et 4 ; Alicke et al., 2008, 2011 ; Lagnado et Channon, 2008 ; Rogers et al., 2019). La raison à cela est qu'Alicke (1992) considère que « de nombreux problèmes de causalité requièrent une évaluation du degré de force causale plutôt qu'un jugement sur la présence ou l'absence d'un facteur particulier » (p.368). En reprenant la logique de Hesslow (1988) et Hilton (1991), nous pourrions considérer que les questions ouvertes devraient principalement permettre aux observateurs de faire une sélection causale (i.e., parmi les conditions présentes, la sélection de la plus pertinente à rapporter dans une conversation), alors que les questions fermées à échelle de réponse leur permettraient davantage d'estimer la force d'un lien causal entre une condition spécifique et le dommage. Ainsi, il est possible que les participants aient indiqué la condition physique (e.g., l'allergie) tout en ayant malgré tout surestimé le lien causal entre l'agent blâmable et le dommage. Dans ce cas, des questions à échelle de réponse centrées sur l'agent seraient plus adaptées. Notre réplique partielle de l'effet de la valence morale pourrait donc soit soutenir la rationalité des jugements causaux (c'est-à-dire le respect des deux critères normatifs), soit refléter un choix méthodologique défavorable à la réplique de cet effet (c'est-à-dire question ouverte vs question à échelle de réponse), soit les deux.

L'étude 2 visait à clarifier cette question. Les participants ont été exposés aux mêmes six scénarios, mais la tâche de jugement consistait ici en une question fermée à échelle centrée sur l'agent. Les hypothèses concernant les critères normatifs sont les mêmes que dans l'étude précédente : l'agent serait plus susceptible d'être considéré comme une cause lorsque le lien causal entre ses actions est fort plutôt que faible et ne serait pas considéré comme une cause lorsqu'il n'y a pas de lien causal (*lien préempté*). En ce qui concerne le *processus de validation du blâme*, nous nous attendions à nouveau à observer l'effet de la valence morale lorsque le rôle causal de l'agent était ambigu (*lien faible*). De plus, étant donné que les questions à échelle de réponse centrées sur l'agent sont plus favorables à l'apparition de l'effet de la valence morale que les questions ouvertes, nous nous attendions également à les reproduire dans la condition

de lien causal fort comme dans l'étude 1. Par conséquent, nous nous attendions à ce que cet effet soit neutralisé seulement lorsque l'agent n'était pas une condition nécessaire réelle au dommage (*lien préempté*).

Étude 2 : Effet de la valence morale sur le jugement causal en fonction des critères normatifs de nécessité réelle et de sensibilité avec une question à échelle de réponse

Méthode

Design et participants. Le plan expérimental était similaire à celui de l'étude 1. Les calculs de puissance statistiques a priori ont été basés sur les tailles d'effet de 10 études³², décrites dans cinq publications différentes sur la validation du blâme, et qui utilisaient des questions à échelle de réponse centrées sur l'agent (Alicke, 1992 ; Alicke et al., 2008, 2011 ; Lagnado & Channon, 2008 ; Rogers et al., 2019). Nous avons obtenu un f de Cohen moyen de 0.464 (effet large, Cohen, 1988, p.285). Le test a indiqué que les données de 19 participants par condition expérimentale permettraient d'atteindre une puissance de 80% avec $\alpha = .05$.

Notre échantillon final était composé de 156 participants (126 femmes), avec un âge moyen de 27.3 ans ($SD = 9.37$; min 18 ; max 60). Notre ensemble de données reflétait une distribution aléatoire de 24 participants dans chaque condition expérimentale. Tous les participants ont été recrutés sur le réseau social Facebook. L'étude a été approuvée par le Comité d'Éthique de la Recherche de Toulouse, et les participants ont donné leur consentement éclairé avant de participer.

Matériel et procédure. Le matériel et la procédure étaient identiques à ceux de l'Étude 1. La seule différence résidait dans la nature de la tâche de jugement causal. Après avoir lu l'un des six scénarios, les participants répondaient à la question suivante : « Dans quelle mesure les

³² Nous avons choisi de ne pas intégrer la taille d'effet de l'étude 2 de Lagnado et Channon (2008) car celle-ci était considérablement plus large que les autres et aurait biaisé la moyenne ($f = 2.140$ pour $N = 80$).

actions et les décisions de Barbara sont-elles la cause de la mort de Jean ? »³³ avec une échelle de réponse³⁴ allant de 1 (pas du tout la cause) à 21 (totalement la cause). Pour mesurer les réactions évaluatives induites par la manipulation des motivations, nous avons demandé aux participants de répondre à cette question : « Comment évalueriez-vous la décision de Barbara de vouloir tuer Jean ? », sur une échelle allant de 1 (pas du tout blâmable) à 21 (totalement blâmable). Un exemple de questionnaire est disponible en Annexe B.

Un test U de Mann-Withney a montré que les participants affectés à la condition *motivations blâmables* ont évalué Barbara plus négativement ($M = 19.73$, $Mdn = 21$, $SD = 4.45$) que les participants affectés à la condition *motivations louables* ($M = 9.62$, $Mdn = 10$, $SD = 5.84$), $U = 451$, $p < .001$. Les distributions des scores d'évaluation en fonction des motivations de l'agent sont disponibles en Annexe C.

Résultats

L'analyse visuelle de nos distributions des scores de jugement causal indiquent que la plupart de nos données ne suivaient pas une loi normale (voir Figure 7, p.89). En effet, les distributions présentaient une asymétrie à gauche dans les conditions du lien fort et lien faible (motivations blâmables), et une asymétrie à droite dans les conditions du lien absent. En revanche, le test de Shapiro-Wilk indiquait que la distribution suivait une loi normale dans la condition lien faible avec un agent ayant des motivations louables ($W = 0.951$, $p = .242$). L'ANOVA serait robuste aux violations de la normalité pour les asymétries (voir Blanca et al., 2017). Son utilisation n'augmenterait pas les erreurs de type 1 (i.e., probabilité de détecter un effet alors que celui-ci n'existe pas). Par conséquent, une ANOVA 2 (Motivations : louables,

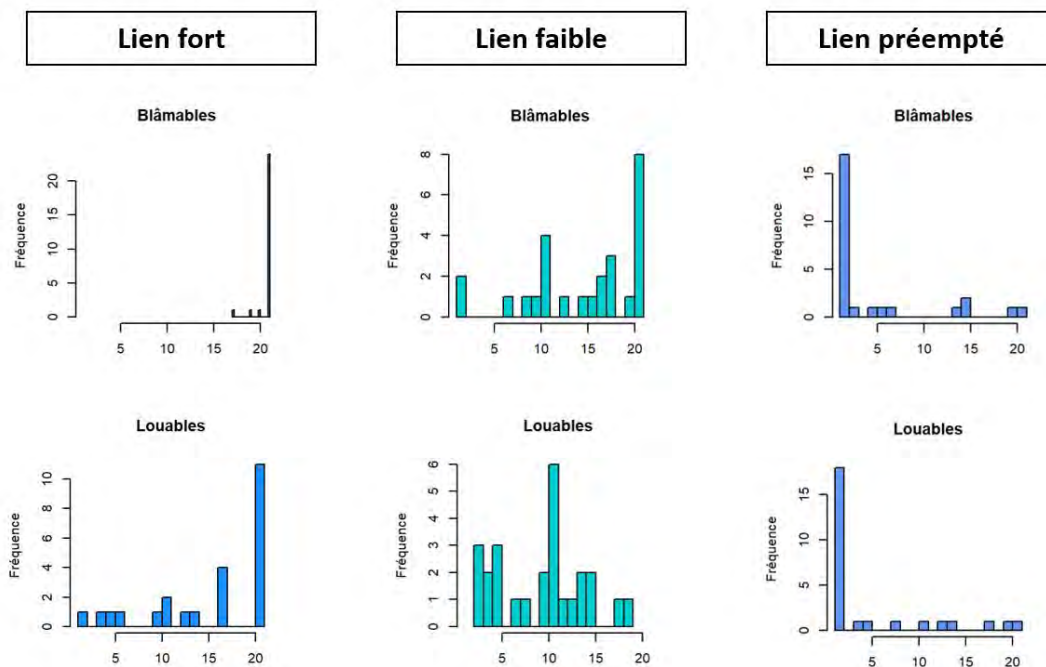
³³ Les résultats de (Hilton et al., 2021, Etude 3) n'ont pas montré de différence significative dans la réplication de l'effet de la valence morale entre les formulations « une cause » et « la cause » pour la question de jugement causal.

³⁴ Nous avons utilisé un curseur plutôt qu'une échelle de Likert classique afin de pouvoir considérer nos variables dépendantes comme d'intervalle (voir Sullivan & Artino, 2013).

blâmables) × 3 (Lien : fort, faible, préempté) a été conduite afin de tester notre hypothèse d'interaction.

Figure 7

Distributions des scores de jugement causal en fonction des motivations de l'agent et du lien causal.



L'analyse a révélé un effet principal significatif du lien causal sur le jugement causal avec une très large taille d'effet, $F(2, 150) = 74.13, p < .001, \eta_p^2 = .497$. D'autres investigations utilisant des contrastes planifiés (contrastés de Helmert) ont révélé que les participants assignés à la condition du lien causal fort ont évalué l'implication causale de Barbara de manière significativement plus élevée ($M = 18.31, Mdn = 21, SD = 5.14$), que les deux autres conditions (lien faible et lien préempté), $t(150) = 10.02, p < .001$. De plus, ces scores de jugement causal étaient significativement plus élevés dans la condition de lien faible ($M = 12.37, Mdn = 11, SD = 6.10$) que dans la condition de lien préempté ($M = 4.98, Mdn = 2, SD = 6.37$), $t(150) = 6.83, p < .001$. Les analyses ont également révélé un effet principal des motivations, avec des scores significativement plus élevés lorsque les motivations étaient blâmables ($M = 13.65, Mdn = 18,$

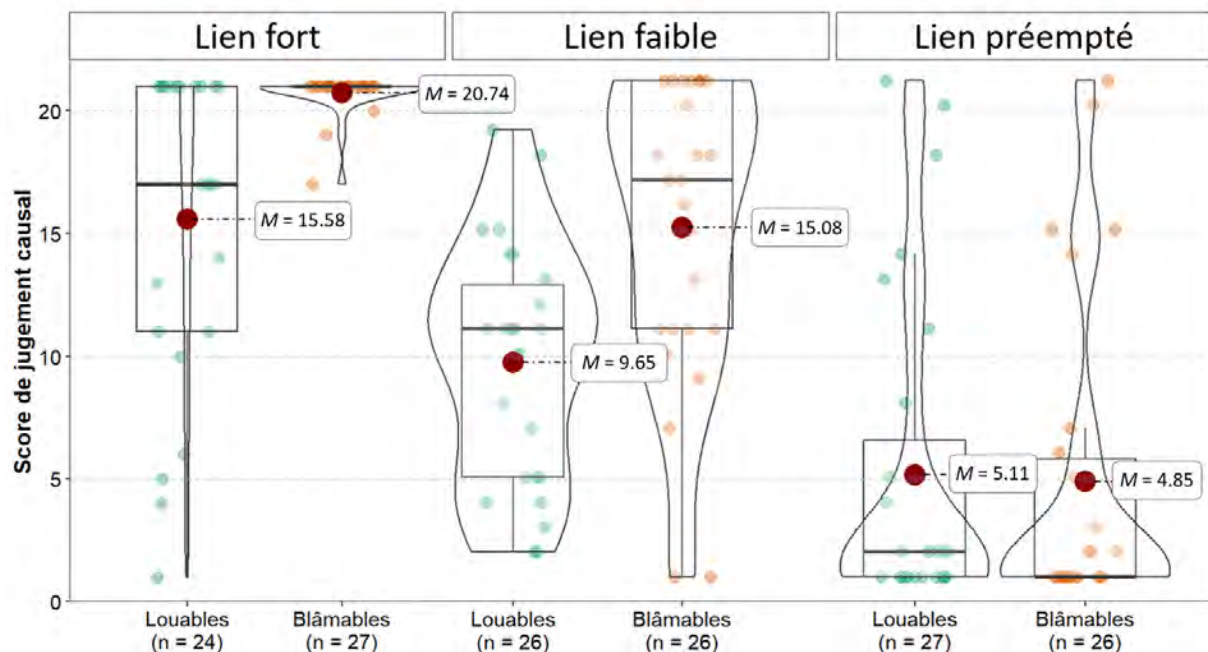
$SD = 8.32$) comparativement à louables ($M = 9.91$, $Mdn = 11$, $SD = 7.28$), $F(1, 150) = 17.73$, $p < .001$, $\eta_p^2 = .089$ (effet moyen).

L'effet d'interaction entre les motivations et le lien causal était également significatif, $F(2, 150) = 4.40$, $p = .014$, $\eta_p^2 = .055$. La Figure 8 (p.91) illustre les statistiques descriptives pour cette interaction. Nous avons examiné de plus près cet effet d'interaction, à l'aide de tests post hoc avec correction de Bonferroni et des t-tests bayésiens par défaut³⁵. Les résultats ont montré que les participants étaient plus susceptibles de juger l'agent aux motivations blâmables (vs. louables) comme la cause du dommage lorsque le lien entre son action et le résultat était fort ($t(150) = 3.32$, $p = .017$, $d = 0.93$, $ICd\ 95\% = [0.37, 1.49]$, $BF_{10} = 152.65$) ou faible ($t(150) = 3.53$, $p = .008$, $d = 0.98$, $ICd\ 95\% = [0.42, 1.54]$, $BF_{10} = 35.98$). Les résultats convergents (1) des fortes tailles d'effet ($d > 0.8$, Cohen, 1988) dont les intervalles de confiance excluent le 0, (2) des facteurs bayésiens indiquant un soutien décisif ($BF_{10} > 100$, lien causal fort) et très fort ($BF_{10} > 30$, lien causal faible) concernant l'hypothèse d'un effet par rapport à l'hypothèse nulle, et (3) de la petite valeur p , nous concluons à la réplication de l'effet de la valence morale dans ces deux conditions. En revanche, l'effet de la valence morale n'a pas été observé dans le lien causal préempté où, conformément à nos hypothèses, aucune différence significative dans le jugement causal n'a été observée entre l'agent ayant des motivations blâmables et celui ayant des motivations louables, $t(150) = 0.17$, $p = 1.00$, $d = 0.05$, $ICd\ 95\% = [-0.49, 0.59]$, $BF_{10} = 0.28$. La valeur p n'étant pas significative, les intervalles de confiance du d de Cohen n'excluant pas 0, et le facteur bayésien apportant un soutien pour l'hypothèse nulle, nous ne pouvons pas conclure à la réplication de l'effet des motivations sur le jugement causal dans le lien causal préempté.

³⁵ Les t-tests seraient également robustes aux asymétries comme les nôtres (Lim & Lim, 2016).

Figure 8

Scores de jugement causal en fonction des Motivations de l'agent et du Lien causal.



Discussion de l'Étude 2

Cette deuxième étude visait à étendre les résultats observés dans l'étude 1 avec une tâche de jugement causal similaire à la majorité des recherches précédemment effectuées sur le phénomène de validation du blâme, c'est-à-dire l'utilisation d'une question fermée centrée sur l'agent avec échelle de réponse. L'observation la plus frappante qui ressort des analyses est la reproduction de l'effet de la valence morale dans une situation où la contribution causale de l'agent envers le dommage est plus ambiguë (i.e., un lien de causalité faible). Dans cette situation, si l'agent voulait que le dommage se produise et a agi pour qu'il se produise, son action dépendait toutefois d'une autre condition qui, en fin de compte, est à l'origine du dommage tel qu'il s'est produit. Ces résultats vont donc plus loin que ceux observés dans l'Étude 1. En effet, s'ils soutiennent l'émergence de l'effet de la valence morale dans le cas d'un lien causal fort, également observés dans l'Étude 1 avec une question ouverte, les résultats de

l'Étude 2 étendent l'observation de ce phénomène aux situations présentant un lien causal faible entre l'agent et le dommage.

Discussion générale des Études 1 et 2

Deux résultats principaux ressortent particulièrement de ces premières études en ce qui concerne la rationalité des jugements causaux de sens commun. Le premier est que le jugement causal dépendrait du respect des deux critères normatifs, à savoir la *nécessité réelle* (Hilton et al., 2021 ; Icard et al., 2017) et la *sensibilité* (Icard et al., 2017 ; Lewis, 1973 ; Woodward, 2006, 2010). En effet, nous avons observé que l'agent était considéré comme la cause du résultat lorsque son action était une condition nécessaire réelle du dommage (*liens forts et faibles vs lien préempté*). De plus, l'agent était plus susceptible d'être identifié comme la cause réelle et de se voir attribuer un score causal plus élevé lorsque son action était insensible au changement des circonstances de fond (*lien fort*) que lorsqu'elle était sensible à ce changement (*lien faible*).

Le deuxième résultat principal est que, même en présence d'une forte propension à blâmer, les jugements causaux s'écartent relativement peu des critères normatifs. En effet, l'influence des réactions évaluatives négatives sur le jugement causal pourrait être conditionnée par les particularités du lien causal. Le fait que les participants aient jugé qu'un agent blâmable (vs. louable) était davantage la cause d'un dommage uniquement lorsque son action était réellement liée au dommage (conditions *liens fort et faible*) peut indiquer que la grande majorité des participants respectaient le critère de nécessité réelle. En effet, même si l'agent avait l'intention d'obtenir un résultat, qu'il a agi pour produire ce résultat et que le résultat qu'il voulait obtenir s'est produit, l'estimation de sa contribution causale était considérablement réduite lorsque son action n'était pas une condition nécessaire au dommage tel qu'il est réellement survenu (condition *lien préempté*). En outre, nous avons démontré que les profanes semblaient prêter attention au critère de sensibilité. D'une part, contrairement aux résultats précédents

(Alicke, 1992 ; Rogers et al., 2019), l'effet de la valence morale n'a pas été reproduit dans l'Étude 1 lorsque l'action de l'agent était sensible aux circonstances de fond (condition *lien faible*). De plus, les résultats de l'Étude 2 ont montré que l'agent recevait un score de contribution causale plus faible lorsque sa force causale diminuait (*lien faible vs lien fort*). En d'autres termes, lorsque les participants devaient désigner la cause de l'événement (question ouverte, Étude 1), l'effet de la valence morale n'était pas reproduit lorsque le lien entre l'action de l'agent et le dommage était faible. Cependant, lorsque les participants ont été invités à donner une estimation plus fine de la contribution causale de l'agent au résultat (question fermée centrée sur l'agent, Étude 2), nous avons reproduit l'effet de la valence morale lorsque le lien entre l'action et le résultat était faible. Ce dernier résultat est davantage conforme aux prédictions du CCM et aux études qui ont testé l'effet de la valence morale avec un lien causal faible (Alicke, 1992).

Dans l'ensemble, les deux principaux résultats peuvent indiquer que lors d'un jugement causal, les individus intégreraient les deux critères normatifs dans leur raisonnement et répondraient à la question causale de manière assez rationnelle. L'effet de la valence morale pourrait alors être un phénomène qui ne se produit que lorsque les deux critères précédents sont remplis. De plus, cet effet semble être assez subtil, car les résultats ont montré qu'il apparaissait principalement lorsque la tâche de jugement se concentrait sur l'agent et impliquait une évaluation du degré de contribution causale.

Compte tenu de la conception des deux études, un aspect méthodologique pourrait néanmoins avoir affecté nos résultats, qui doivent être interprétés avec prudence. Premièrement, dans les études précédentes s'inscrivant dans le CCM (Alicke, 1992 ; Alicke et al., 2008, 2011), les chercheurs ont utilisé des scénarios dans lesquels les deux agents (louable vs. blâmable) avaient des états épistémiques différents (i.e., intentionnalité vis-à-vis du

dommage). En effet, dans ces scénarios, l'agent aux motivations louables ne voulait pas que le dommage se produise. Cette opérationnalisation avait l'avantage de présenter un contraste important entre la valence morale de l'agent louable et celle de l'agent blâmable, plus favorable à l'observation de l'effet de la valence morale. En effet, le CCM stipule que la force du lien causal d'un agent blâmable devrait être exagérée si « les évaluations spontanées sont suffisamment fortes » (Alicke et al., 2011, p.9). Dans nos scénarios, l'état épistémique de l'agent était comparable entre les conditions expérimentales louables et blâmables : l'intention du résultat et l'intention d'agir étaient toujours présentes. Cependant, ce choix méthodologique peut entraîner le risque que les participants dans la condition motivations louables (épouse régulièrement maltraitée par son mari) n'aient pas perçu les motivations comme réellement louables. En France, l'attitude du public sur la légitime défense différée n'est pas totalement connue à ce jour. Cette notion, très discutée dans l'affaire judiciaire de Jacqueline Sauvage (2012-2014), permet d'exonérer une personne qui a agi de manière non concomitante à l'agression pour se protéger d'autres agressions à l'avenir (ex : violences conjugales répétées). Bien que les participants aient évalué Barbara moins négativement lorsque ses motivations étaient louables (vs. blâmables), on ne peut pas exclure que le fait de tuer une personne, pour quelque raison que ce soit, reste moralement répréhensible pour nos participants. Cela impliquerait un contraste plus faible entre les conditions Motivations (louables, blâmables) dans nos études actuelles par rapport aux études précédentes. En effet, bien qu'un consensus au niveau de l'évaluation morale de l'agent semble s'être formé parmi les participants dans la condition *motivations blâmables*, il semble moins évident dans la condition *motivations louables* (cf. distributions des scores d'évaluation en Annexe C). Cette limitation souligne un problème méthodologique : Dans une condition de lien causal fort, il est impossible d'élaborer une condition expérimentale véritablement « louable » dans laquelle l'agent souhaitait qu'un mauvais résultat se produise et a agi de manière à produire ce résultat. La comparaison des

conditions de liens causaux faibles et préemptés est la seule option pour proposer un scénario avec un agent louable qui a agi pour produire un résultat positif, mais dont l'action a été déviée ou préemptée pour produire un incident à la fin.

Étude 3 : Effet de la valence morale sur le jugement causal en fonction du critère normatif de nécessité réelle avec un agent aux motivations réellement louables

Les modèles causaux que nous avons utilisés ne permettent pas de mettre en avant un agent réellement louable. En effet, l'agent de la condition *motivations louables* souhaitait provoquer un dommage (e.g., une femme battue qui veut tuer son mari). Par conséquent, le contraste entre les deux conditions vis-à-vis du caractère blâmable de l'agent n'est potentiellement pas assez important pour voir apparaître l'effet de la valence morale. Nous proposons ainsi une nouvelle opérationnalisation de la valence morale de l'agent, en particulier de la condition *motivations louables*. Il s'agit d'un scénario dans lequel l'agent avait l'intention de provoquer un résultat positif (motivations louables) ou négatif (motivations blâmables). Néanmoins, son action a été déviée par une autre condition et a provoqué un dommage³⁶. Ce scénario présente ainsi un agent aux motivations louables davantage comparable aux études s'inscrivant dans l'approche du CCM (Alicke, 1992 ; Alicke et al., 2008, 2011 ; Rogers et al., 2019). Cette opérationnalisation serait alors plus favorable à l'apparition de l'effet de la valence morale, par le biais d'un contraste plus important entre l'évaluation vis-à-vis de l'agent louable et celle vis-à-vis de l'agent blâmable. Néanmoins, cette opérationnalisation ne permettra de tester que le respect du critère de nécessité réelle (la condition lien causal fort ne pouvant plus être mise en place).

³⁶ L'action de l'agent étant déviée dans les deux conditions, ces scénarios sont comparables à la condition lien faible de nos deux premières études.

L'objectif principal de cette étude était de tester notre hypothèse d'interaction entre la valence morale et le critère de nécessité réelle, avec une opérationnalisation qui serait considérablement propice à l'activation du *processus de validation du blâme*. Malgré la présence d'un contraste plus important entre les conditions de valence morale, nos hypothèses restent les mêmes : (1) l'agent serait considéré comme cause uniquement lorsqu'il est une condition nécessaire réelle au dommage, et (2) l'effet de la valence morale surviendrait uniquement s'il existe un lien causal entre l'action de l'agent et le dommage (*lien présent vs absent*)³⁷.

Nous avons également souhaité approfondir l'étude des facteurs déterminants l'évaluation négative vis-à-vis de l'agent. Pour ce faire, nous avons affiné la mesure de l'évaluation des participants vis-à-vis de l'agent en mesurant l'évaluation globale de l'agent (i.e., sa personne), ses intentions ainsi que ses actions. Comme d'autres auteurs (Alicke et al., 2011 ; Rogers et al., 2019), nous avons également souhaité tester l'hypothèse de médiation du CCM, à savoir que l'évaluation vis-à-vis de l'agent serait une variable médiatrice (Baron & Kenny, 1986) de l'effet des motivations de l'agent sur le jugement causal. Les différentes mesures de l'évaluation nous permettront d'apporter davantage de précisions sur cet effet de médiation identifié par les auteurs précédemment cités. Notre hypothèse est que les trois variables de l'évaluation (personne, intentions, actions) médieront l'effet des motivations sur le jugement causal, mais uniquement dans la condition d'un lien causal présent.

Enfin, au regard des applications sociétales de nos recherches dans le domaine judiciaire, nous avons également souhaité étudier les attributions de responsabilité et de sanctions envers l'agent. Nos hypothèses concernant ces attributions étaient identiques à celles

³⁷ La condition *lien présent* est comparable à la condition *lien faible* dans les études 1 et 2 ; La condition *lien absent* est comparable à la condition *lien préempté* dans les études 1 et 2.

vis-à-vis du jugement causal. En effet, concernant l'attribution d'une responsabilité, la présence d'un lien causal entre l'agent et le dommage doit être établie avant de prendre en considération la présence d'états mentaux blâmables (Bordel et al., 2006 ; Fincham & Jaspars, 1980 ; Hamilton, 1978 ; Lagnado & Gerstenberg, 2017). Il en irait de même pour l'attribution de sanctions (Cushman, 2008 ; Fincham & Shultz, 1981 ; Shultz et al., 1981, 1986). Par conséquent, nous nous attendions à des attributions de responsabilité et de sanctions réduites et à aucun effet des motivations de l'agent lorsque l'agent n'est pas une condition nécessaire au dommage. Il convient d'indiquer que la présence d'un effet de la valence morale sur ces attributions en présence d'un lien causal ne constitue pas un biais, contrairement au jugement causal (cf. p.24 sur l'attribution de responsabilité).

Méthode

Design. Le plan expérimental était un plan factoriel 2 (Motivations : louables, blâmables) × 2 (Lien causal : présent, absent) en inter-sujets.

Participants. Conformément aux analyses de puissance de l'étude précédente, nous visions un minimum de 19 participants par condition expérimentale pour atteindre une puissance statistique de 80% avec $\alpha = .05$. Recrutés par le biais d'une annonce sur le réseau social Facebook, 160 personnes ont pris part à cette étude (soit 20 participants par condition), dont 122 femmes (76.2%) et 38 hommes (23.8%). La moyenne d'âge était de 25.5 ans ($SD = 10.26$). Tous les participants étaient des volontaires et n'ont reçu aucune compensation financière pour leur participation.

Matériel et procédure. La passation de l'étude se déroulait en ligne par l'intermédiaire du logiciel Qualtrics. En premier lieu, les participants étaient invités à lire le formulaire de consentement et à indiquer leur accord s'ils l'approuvaient. Ils renseignaient ensuite des informations démographiques (genre et âge), puis il leur était demandé de lire attentivement

l'un des 8 scénarios distribués aléatoirement. Immédiatement après la lecture, ils étaient invités à répondre à une série de 6 questions sur l'évaluation de l'agent, le jugement causal et l'attribution de responsabilité et de sanctions.

Scénarios. Deux scénarios différents ont été créés pour contrôler l'effet du type d'événement sur les jugements. Dans chacun de ces scénarios (feu de forêt/propagation de la salmonelle), les motivations de l'agent étaient louables ou blâmables pour manipuler les réactions évaluatives telles que décrites dans le *processus de validation du blâme*. Les scénarios sont disponibles en Annexe D.

Le premier scénario décrivait un feu de forêt, dans lequel David allumait un feu pour contrer un feu sauvage qui menaçait un village (motivations louables) vs. pour mener le feu sauvage au village (motivations blâmables). Dans la condition *motivations louables*, l'agent était présenté comme un garde forestier ayant pris toutes les précautions nécessaires pour allumer un contre-feu, de sorte qu'il ne puisse être accusé de négligence car un acte négligent est évalué plus négativement qu'un acte accidentel (e.g., Karlovac & Darley, 1988 ; Shultz & Wright, 1985). Dans la condition *motivations blâmables*, l'agent était présenté comme un pyromane.

Dans le second scénario dit de la Salmonelle, Quentin, l'employé d'un restaurant, baissait la température de la chambre froide pour diminuer les risques de prolifération de la salmonelle (*motivations louables*) vs. pour les augmenter (*motivations blâmables*).

Ces 4 grands scénarios étaient ensuite structurés en deux chaînes causales : l'une présentant un lien causal (l'action de l'agent est une condition nécessaire réelle à la survenue du résultat), l'une ne présentant pas de lien causal (l'action de l'agent est non nécessaire à la survenue du résultat).

Huit scénarios ont donc été construits, soit quatre pour chaque scénario (feu de forêt et salmonelle).

Lien causal présent. Pour les scénarios du feu de forêt, le premier feu s'éteignit et une brise non prévue par l'agent (garde-forestier, pyromane) mena son feu jusqu'au village et tua douze personnes. Dans les scénarii de la salmonelle, le thermostat avait subi un dysfonctionnement, l'agent augmenta donc la température au lieu de la baisser (condition *motivations louables*) vs. l'agent la baissa au lieu de l'augmenter et le refroidissement intense provoqua une panne du système de refroidissement (condition *motivations blâmables*), ceci permit à la salmonelle de proliférer et des clients tombèrent malades.

Lien causal absent. Dans le scénario du feu de forêt, le feu sauvage et le feu provoqué par l'agent (motivations : louables, blâmables) s'éteignirent. Soudainement, la foudre vint frapper un arbre qui créa un feu se dirigeant vers le village et tua douze personnes. Dans le scénario de la salmonelle, l'action de l'agent (motivations : louables, blâmables) sur le thermostat n'affecta pas la température en raison d'un dysfonctionnement. Soudainement, un court-circuit indépendant de l'action de l'agent provoqua un réchauffement de la chambre froide qui engendra les mêmes dommages que dans la condition lien causal présent.

Questionnaire. Après avoir lu le scénario, les participants répondaient à six questions. Trois questions visaient à mesurer les réactions évaluatives des participants. Les trois autres questions, quant à elles, mesuraient le jugement causal, les attributions de responsabilité et de sanctions, respectivement.

Les participants répondaient aux questions de réactions évaluatives sur une échelle de -10 (totalement blâmable) à 10 (totalement louable). Plus précisément, ils évaluaient l'agent lui-même (e.g., Comment évaluez-vous David ?), ses intentions (e.g., Comment évaluez-vous

les intentions de David envers le village ?) et son action (e.g., Comment évaluez-vous le fait que David ait mis délibérément le feu à une zone de la forêt ?). En plus de tester les hypothèses présentées en introduction, ces trois mesures permettaient également de s'assurer que les motivations de l'agent étaient correctement opérationnalisées, en particulier les motivations louables de l'agent qui provoque un résultat négatif.

Notre question visant à mesurer le jugement causal était la suivante : « Dans quelle mesure les actions de l'agent ont-elles causé le dommage ? » avec une échelle de 1 (pas du tout causé) à 21 (totalement causé). Deux questions ont également été posées pour étudier l'effet des motivations et du lien causal sur l'attribution de responsabilité et de proposition de sanctions : « Dans quelle mesure tenez-vous [nom de l'agent correspondant] pour responsable de la mort des habitants ? » et « Dans quelle mesure êtes-vous d'accord que [nom de l'agent correspondant] soit sanctionné ? ».

Résultats

Évaluation de l'agent. Afin de contrôler au mieux cette nouvelle opérationnalisation de la valence morale de l'agent, nous avons effectué un examen approfondi des distributions de chaque dimension de l'évaluation (personne, intentions, actions).

Dans le scénario du Feu de forêt, les distributions des scores d'évaluation étaient caractérisées par une forte asymétrie à gauche lorsque l'agent avait des motivations louables et une forte asymétrie à droite lorsque ses motivations étaient blâmables (voir Figure 9, p.101). En revanche, les distributions présentaient une variance élevée sur les scores d'évaluation de sa personne et de ses actions, lorsque les motivations de l'agent du scénario de la Salmonelle étaient louables (voir Figure 10, p.102). Ce constat sur les distributions implique que l'opérationnalisation des motivations de l'agent dans le scénario du Feu de forêt serait plus valide que dans le scénario de la Salmonelle. En effet, l'examen du Tableau 3 (voir p.102)

indique qu'il n'y avait pas de consensus parmi les participants du scénario de la Salmonelle pour évaluer l'agent louable positivement sur le plan de sa personne et de ses actions. La médiane de ces deux dimensions, mesurées sur une échelle allant de -10 à 10, était effectivement proche de 1 (le 0 signifie une évaluation neutre). Il conviendra de considérer cette différence entre les scénarios lors de l'analyse des résultats. Malgré ce constat, les tests *U* de Mann-Whitney ont montré que la différence entre l'agent louable et blâmable était statistiquement significative pour chaque dimension de l'évaluation et ce, quel que soit le scénario (voir Tableau 3, p.102).

Figure 9

Distributions des scores d'évaluation de la personne, de ses intentions et de ses actions en fonction des motivations de l'agent pour le scénario du Feu de Forêt.

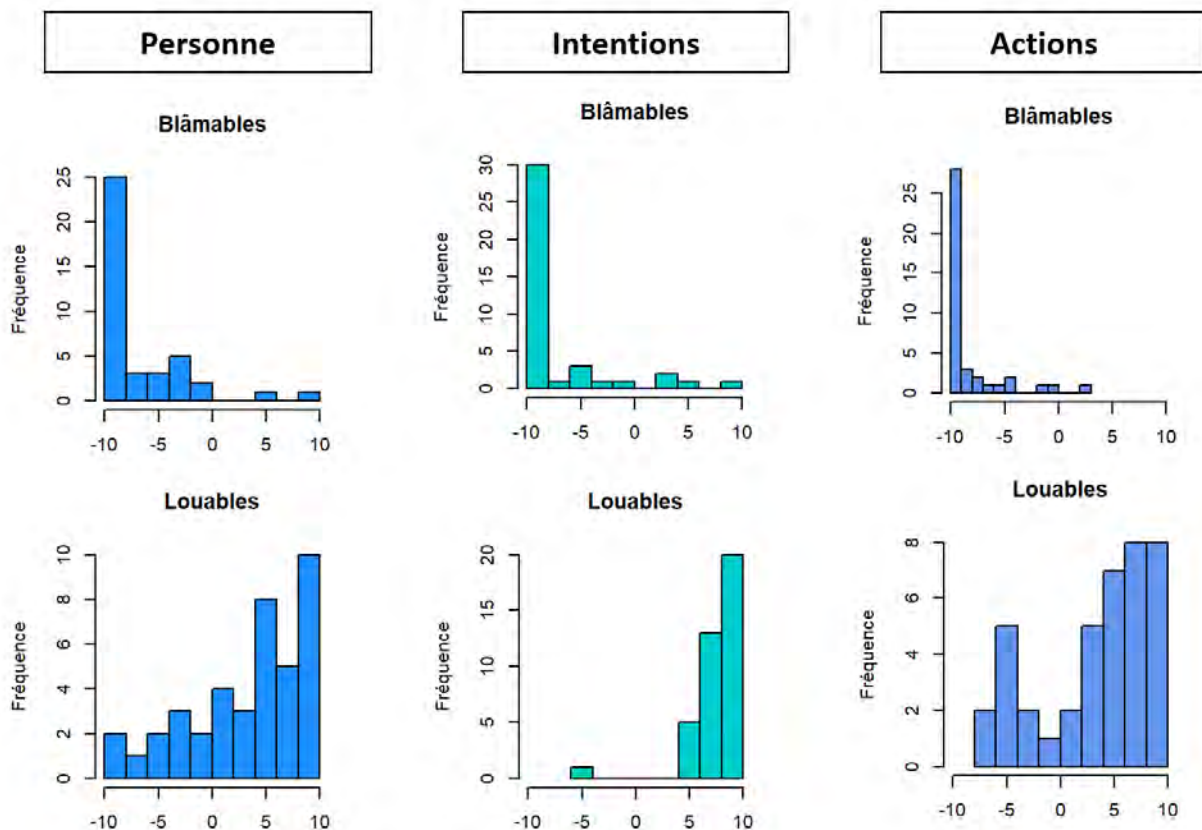


Figure 10

Distributions des scores d'évaluation de la personne, de ses intentions et de ses actions en fonction des motivations de l'agent pour le scénario de la Salmonelle.

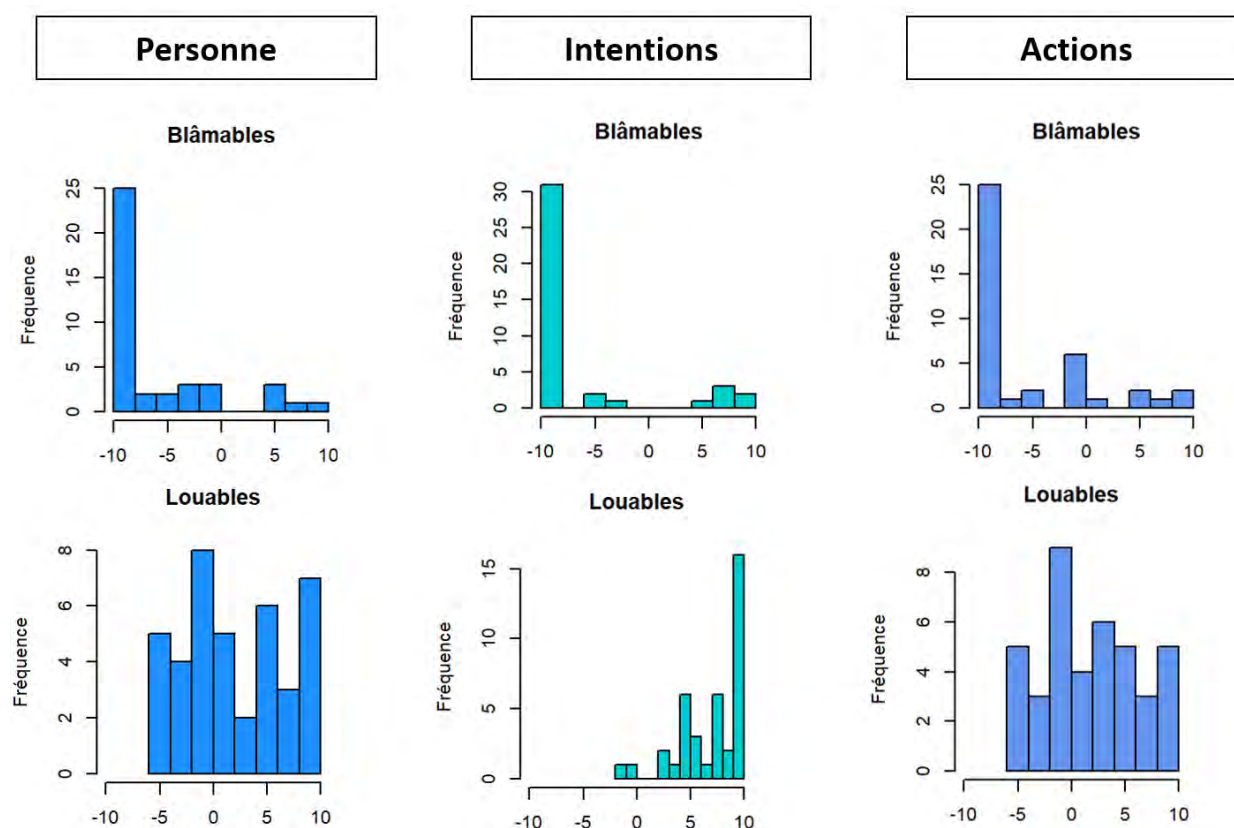


Tableau 3

Scores médians d'évaluation de l'agent (sa personne, ses intentions et ses actions) en fonction de ses motivations et du scénario

	Feu de forêt			Salmonelle		
	Louables (n = 40)	Blâmables (n = 40)	U	Louables (n = 40)	Blâmables (n = 40)	U
Personne	5.50 [0.75, 8.25]	-9.50 [-10.00, -5.00]	132.0***	1.00 [-0.25, 6.50]	-10.00 [-10.00, -3.00]	218***
Intentions	9.00 [7.00, 10.00]	-10.00 [-10.00, -7.50]	37.0***	8.00 [5.00, 10.00]	-10.00 [-10.00, -8.00]	121***
Actions	5.00 [1.50, 7.00]	-10.00 [-10.00, -8.00]	51.5***	1.50 [-0.25, 5.00]	-10.00 [-10.00, 0.00]	276***

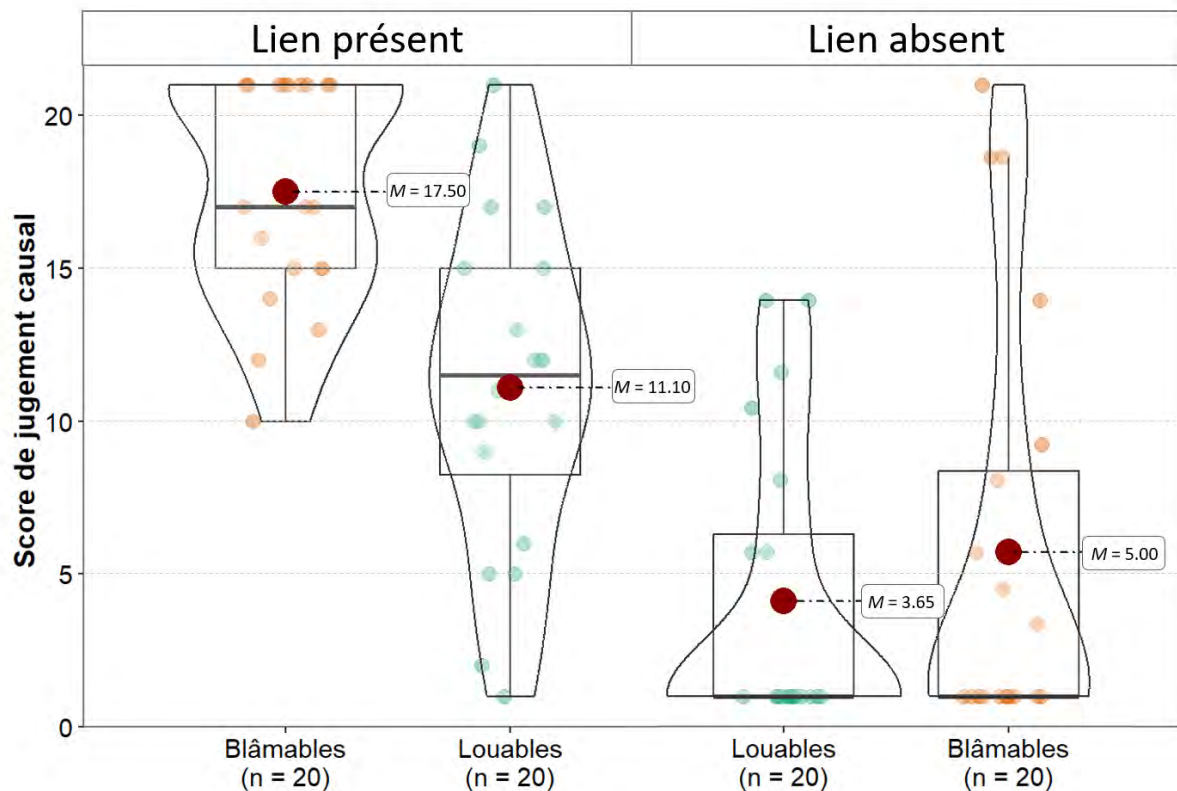
Note. Les quartiles sont entre crochets [Q1, Q3]. *** $p < .001$

Jugement causal. Afin de tester notre hypothèse concernant l'interaction entre l'effet de la valence morale et le critère de nécessité réelle, une ANOVA 2 (lien causal : présent, absent) \times 2 (motivations : louables, blâmables) a été conduite sur le score de jugement causal pour chaque scénario (correction de l'alpha à .025).

Scénario du feu de forêt. L'agent se voyait attribuer un score de causalité plus élevé dans la condition du *lien présent* ($M = 14.30$, $Mdn = 15.00$, $SD = 5.59$) que dans la condition du *lien absent* ($M = 4.32$, $Mdn = 1.00$, $SD = 5.03$), $F(1, 76) = 85.00$, $p < .001$, $\eta_p^2 = .528$. Comme dans les études précédentes, la large taille d'effet illustre l'importance de cette variable sur les jugements causaux. Les analyses ont également révélé un effet principal des motivations de l'agent : les participants attribuaient un score plus élevé à l'agent blâmable ($M = 11.25$, $Mdn = 13.50$, $SD = 7.95$) qu'à l'agent louable ($M = 7.38$, $Mdn = 6.50$, $SD = 6.05$), $F(1, 76) = 12.83$, $p < .001$, $\eta_p^2 = .144$. Enfin, comme l'illustre la Figure 11 (p.104), une interaction statistiquement significative a été observée entre le lien causal et les motivations de l'agent, $F(1, 76) = 5.45$, $p = .022$, $\eta_p^2 = .067$. Plus précisément, lorsqu'un lien causal unissait les actions de l'agent au dommage, l'effet des motivations sur le score de jugement causal était répliqué, $t(76) = 4.18$, $p < .001$, $d = 1.32$, $ICd\ 95\% = [0.66, 1.99]$, $BF_{10} = 237.58$. La forte taille d'effet, dont les intervalles de confiance excluaient le 0, le facteur bayésien indiquant que les résultats sont 237.58 fois plus probables sous l'hypothèse d'un effet que sous l'hypothèse nulle, et la petite valeur p soutenaient la réplification de l'effet de la valence morale dans cette condition. En revanche, tous les résultats convergeaient vers l'absence d'un effet des motivations sur le jugement causal lorsque l'agent n'était pas une condition nécessaire réelle au dommage, $t(76) = 0.88$, $p = 1.00$, $d = 0.28$, $ICd\ 95\% = [-0.91, 0.35]$, $BF_{10} = 0.41$.

Figure 11

Scores de jugement causal en fonction des Motivations de l'agent et du Lien causal pour le scénario du Feu de forêt.



Scénario de la salmonelle. Les analyses ont montré que l'employé dont les actions étaient causalement liées à l'intoxication des clients se voyait davantage désigné comme cause ($M = 15.28$, $Mdn = 17$, $SD = 6.01$), que l'agent dont le lien action-dommage n'était pas présent ($M = 8.10$, $Mdn = 6.00$, $SD = 7.38$), $F(1, 76) = 23.56$, $p < .001$, $\eta_p^2 = .237$. Bien qu'importante, la taille d'effet était cependant plus faible pour ce scénario que pour celui du feu de forêt. Nous n'observons pas d'effet significatif des motivations sur le jugement causal, $F(1, 76) = 0.80$, $p = .373$, $\eta_p^2 = .010$. Par ailleurs, la valeur p de l'effet d'interaction était supérieure à notre seuil alpha de .025, $F(1, 76) = 4.05$, $p = .048$, $\eta_p^2 = .051$. En raison d'une distribution anormale bimodale ($W = 0.905$, $p = .003$) dans la condition *lien absent* et *motivations louables*, nous avons mené des tests non-paramétriques U de Mann-Whitney pour les tests post-hoc. L'effet

des motivations sur le jugement causal n'a été répliqué ni dans la condition du lien présent ($U = 149, p = .164$), ni dans celle du lien absent, ($U = 132, p = .057$).

Analyses de médiation. Conformément aux recommandations (Hayes & Scharkow, 2013 ; Preacher & Hayes, 2004 ; Zhao et al., 2010), nous avons utilisé une méthode non paramétrique par bootstrap³⁸ dans la condition ayant répliqué l'effet de la valence morale (i.e., Lien présent dans le scénario du feu de forêt) pour tester cette hypothèse. Les résultats de ces analyses sont disponibles dans le Tableau 4 ci-après. Un effet direct qualifie l'effet de la VI sur la VD (Motivations → Jugement causal) en contrôlant la variable médiatrice (Évaluation). Un effet indirect renvoie à l'effet de médiation (Motivations → Évaluation → Jugement causal). Enfin, un effet total correspond à l'effet de la VI sur la VD en excluant la variable médiatrice. Les résultats indiquaient que nous n'avions aucun effet indirect significatif, ce qui signifierait que la relation entre les motivations et le jugement causal ne s'explique pas par les différentes mesures d'évaluation.

Tableau 4

Analyses de médiation par bootstrap pour chaque dimension de l'évaluation de l'agent ainsi que son évaluation générale sur le score de jugement causal pour la condition du lien causal présent du scénario du feu de forêt

Méiateur	Effet	Estimation	Intervalle de confiance (95%) [inférieur, supérieur]	<i>p</i>
Personne	Direct	5.39	[0.11, 9.14]	.048*
	Indirect	1.01	[-0.80, 5.62]	.324
	Total	6.40	[3.55, 9.29]	< .001*
Intentions	Direct	3.94	[-1.21 ; 12.83]	.113
	Indirect	2.40	[-5.22 ; 6.81]	.383
	Total	6.34	[3.35 ; 9.36]	< .001*
Actions	Direct	5.51	[0.42 ; 10.39]	.035*
	Indirect	0.89	[-2.56 ; 4.83]	.590
	Total	6.40	[3.55 ; 9.19]	< .001*
Générale	Direct	3.98	[-2.28 ; 9.50]	.185
	Indirect	2.36	[-2.11 ; 7.61]	.269
	Total	6.34	[3.42 ; 9.25]	< .001*

³⁸ Nous avons effectué 5000 itérations comme dans l'article de Rogers et al (2019)

Jugements de responsabilité et de sanctions. Enfin, nous avons étudié l'impact du lien causal et des motivations de l'agent sur les jugements de responsabilité et de sanctions. Pour ce faire, une ANOVA 2 (Lien causal : présent, absent) × 2 (Motivations : louables, blâmables) a été conduite sur ces scores pour chaque scénario (correction de l'alpha à .025).

Tableau 5

Scores moyens de jugement de responsabilité et de sanctions en fonction du lien causal et des motivations de l'agent pour le scénario du Feu de forêt

Jugement	Lien causal	Motivations		<i>t</i> (<i>df</i>)	<i>p</i>	<i>d</i> [95% IC]	BF ₁₀
		Louables	Blâmables				
Responsabilité	Présent	9.45 (4.27)	16.80 (4.16)	5.12 (76)	< .001	1.62 [0.94, 2.30]	5188.82
	Absent	2.50 (3.35)	5.20 (5.96)	1.88 (76)	.382	0.56 [-0.04, 1.23]	1.04
Sanctions	Présent	6.20 (4.86)	17.30 (3.76)	6.72 (76)	< .001	2.12 [1.41, 2.84]	7.91 × 10 ⁶
	Absent	1.85 (3.22)	9.30 (7.82)	4.51 (76)	< .001	1.43 [0.75, 2.10]	76.97

Note. Les écarts-types sont entre parenthèses

Scénario du feu de forêt. Les statistiques descriptives et les tests post-hoc sont présentés dans le Tableau 5 ci-dessus. L'agent était davantage jugé responsable ($M = 13.12$, $SD = 5.58$) et sanctionné ($M = 11.75$, $SD = 7.07$) en présence d'un lien causal qu'en l'absence de celui-ci ($M_{responsabilité} = 3.85$, $SD_{responsabilité} = 4.96$; $M_{sanctions} = 5.58$, $SD_{sanctions} = 7.01$) avec respectivement : $F(1, 76) = 83.59$, $p < .001$, $\eta_p^2 = .524$ et $F(1, 76) = 27.92$, $p < .001$, $\eta_p^2 = .269$. Le même pattern de résultats a été observé pour l'effet principal des motivations, l'agent blâmable était davantage considéré comme responsable ($M = 11.00$, $SD = 7.76$) et sanctionné ($M = 13.30$, $SD = 7.29$) que l'agent louable ($M_{responsabilité} = 5.97$, $SD_{responsabilité} = 5.17$; $M_{sanctions} = 4.03$, $SD_{sanctions} = 4.63$), avec respectivement : $F(1, 76) = 24.54$, $p < .001$, $\eta_p^2 = .244$ et $F(1,$

76) = 62.98, $p < .001$, $\eta_p^2 = .453$. À l'inverse du jugement de responsabilité, la taille d'effet des motivations était plus importante que celle du lien causal concernant l'attribution des sanctions. Le facteur Lien causal contribuait majoritairement au jugement de responsabilité, alors que le facteur Motivations participait le plus à l'attribution de sanctions.

Enfin, nous observions un effet d'interaction entre les motivations et le lien causal statistiquement significatif pour les jugements de responsabilité, $F(1, 76) = 5.25$, $p = .024$, $\eta_p^2 = .065$. Cependant, nos résultats ne montraient aucune interaction significative pour l'attribution de sanctions, $F(1, 76) = 2.44$, $p = .123$, $\eta_p^2 = .031$. Comme indiqué dans le Tableau 5 (p.106), l'agent blâmable (vs. louable) était davantage désigné comme responsable mais uniquement dans la condition d'un *lien présent*. En revanche, l'agent blâmable était davantage sanctionné par rapport à l'agent louable avec de larges tailles d'effet, et ce, indépendamment de la présence d'un lien causal entre ses actions et le dommage.

Tableau 6

Scores moyens de jugement de responsabilité et de sanctions en fonction du lien causal et des motivations de l'agent pour le scénario de la Salmonelle

	Lien causal	Motivations		<i>t</i> (<i>df</i>)	<i>p</i>	<i>d</i> [95% IC]	BF ₁₀
		Louables	Blâmables				
Responsabilité	Présent	10.65 (4.98)	16.15 (6.66)	2.58 (76)	.071	0.82 [0.17, 1.46]	8.14
	Absent	10.20 (6.34)	6.85 (8.52)	1.57 (76)	.722	0.50 [-1.13, 0.14]	0.68
Sanctions	Présent	6.55 (4.42)	15.20 (7.13)	4.22 (76)	< .001	1.33 [0.67, 2.00]	434.30
	Absent	6.50 (6.52)	12.25 (7.43)	2.80 (76)	.038	0.89 [0.24, 1.53]	4.02

Note. Les écarts-types sont entre parenthèses

Scénario de la salmonelle. Les statistiques descriptives et les tests post-hoc sont présentés dans le Tableau 6 (p.107). Dans ce scénario, nos résultats montraient que l'employé était davantage jugé responsable dans la condition d'un lien causal présent ($M = 13.40$, $SD = 6.44$), que dans la condition d'un lien causal absent ($M = 8.53$, $SD = 7.60$), $F(1, 76) = 10.46$, $p = .002$, $\eta_p^2 = .121$. En revanche, l'employé blâmable ($M = 10.43$, $SD = 5.63$) n'était pas considéré comme plus responsable de l'incident que l'employé louable ($M = 11.50$, $SD = 8.90$), $F(1, 76) = 0.51$, $p = .478$, $\eta_p^2 = .007$.

Nos analyses ont montré que l'effet d'interaction entre le lien causal et les motivations était statistiquement significatif, $F(1, 76) = 8.62$, $p = .004$, $\eta_p^2 = .102$. Comme indiqué dans le tableau 6, l'effet des motivations sur l'attribution de responsabilité dans la condition d'un lien causal présent affichait une valeur p supérieure au seuil alpha. Néanmoins, les intervalles de confiance de la taille d'effet excluaient le 0 et le facteur bayésien apportait un soutien modéré pour l'hypothèse alternative. Cependant, lorsque le lien causal était absent, l'hypothèse alternative ne bénéficiait d'aucun soutien.

Concernant l'attribution de sanctions, l'employé en recevait autant quel que soit le lien causal : présent ($M = 10.88$, $SD = 7.31$) ou absent ($M = 9.38$, $SD = 7.49$), $F(1, 76) = 1.07$, $p = .304$, $\eta_p^2 = .014$. Néanmoins, il était davantage sanctionné lorsque ses motivations étaient blâmables ($M = 13.72$, $SD = 7.34$) plutôt que louables ($M = 6.53$, $SD = 5.50$), $F(1, 76) = 24.66$, $p < .001$, $\eta_p^2 = .245$. Ce résultat allait dans le sens de ceux du scénario du feu de forêt en indiquant que l'attribution de responsabilité repose sur le lien causal et que l'attribution de sanctions est principalement impactée par les motivations de l'agent.

Enfin, les analyses n'ont pas montré d'effet d'interaction entre le lien causal et les motivations pour l'attribution de sanctions, $F(1, 76) = 1.00$, $p = .320$, $\eta_p^2 = .013$. En effet, les

motivations avaient un effet sur l'attribution de sanctions indépendamment du lien causal. Les résultats semblaient toutefois donner davantage de poids à l'effet dans la condition du lien causal présent : valeur p très faible, grande taille d'effet et facteur bayésien décisif pour l'hypothèse alternative (voir Tableau 6). Ce soutien était plus modéré lorsque le lien causal était absent.

Discussion de l'Étude 3

Cette étude avait pour objectif de poursuivre l'examen de l'articulation entre le critère de nécessité réelle et l'effet de la valence morale dans des conditions plus favorables à leur apparition (i.e., par le biais d'une opérationnalisation différente des motivations louables).

Le principal résultat concernant la rationalité des jugements causaux est que, à nouveau, l'effet de la valence morale a été répliqué uniquement lorsque les actions de l'agent constituaient une condition nécessaire réelle à la survenue du dommage. Ces résultats ont également été étendus à l'attribution de la responsabilité. Ceci corrobore de précédents résultats (Fincham & Jaspars, 1983 ; Fincham & Roberts, 1985 ; Shultz et al., 1981, 1986 ; Shultz & Wright, 1985) soutenant que le jugement de responsabilité de sens commun se déroulerait de manière séquentielle et rationnelle : l'individu établirait d'abord la présence d'un lien causal entre l'agent et le dommage, avant de considérer si ses états mentaux sont répréhensibles. Si l'agent n'est pas une condition nécessaire réelle, alors ses états mentaux ne sont pas pris en compte dans l'attribution de responsabilité. C'est précisément ce qui semble s'être passé ici dans nos deux scénarios.

A l'inverse, les motivations de l'agent importaient davantage que le lien causal pour l'attribution de sanctions. En effet, dans le scénario du Feu de forêt, la taille d'effet des motivations était plus grande que celle du lien causal. De plus, les motivations avaient un effet sur l'attribution de sanctions, indépendamment de la présence d'un lien causal. Ce résultat est

encore plus marquant dans le scénario de la Salmonelle, dans lequel seules les motivations avaient un effet sur l'attribution de sanctions (l'effet du lien causal n'était pas statistiquement significatif). Ces résultats contredisent d'autres résultats de la littérature soutenant que, comme l'attribution de responsabilité, un lien causal doit être présent avant d'évaluer les états mentaux de l'agent (Cushman, 2008 ; Fincham & Shultz, 1981 ; Shultz et al., 1981, 1986).

Une différence notable entre ces études et la nôtre pourrait être la sévérité du dommage. En effet, nous avons utilisé un scénario propice aux effets de validation du blâme, à savoir le décès de plusieurs personnes (scénario Feu de forêt) et un empoisonnement massif (scénario Salmonelle). Dans les études précédemment citées, le dommage était moins important (e.g., une alliance perdue, une main brûlée, une maquette d'avion endommagée). Or la sévérité du dommage aurait tendance à augmenter la colère ressentie par les participants (Malle et al., 2014). Notons tout de même que les études sur la sévérité des conséquences sur le jugement moral et l'attribution de responsabilité sont assez contrastées, certaines n'ayant obtenu aucun effet sur le jugement moral (e.g., Arkkelin, Oakley, & Mynatt, 1979 ; Thomas & Parpal, 1987). Sarin, Lagnado et Burgess (2017) n'ont pas non plus obtenu de changement dans l'attribution de responsabilité lorsqu'ils ont fait varier la sévérité des conséquences. Néanmoins, les conséquences négatives dans les scénarii de cette dernière étude étaient d'une faible gravité comme par exemple un jardin rendu infertile, une robe mal cousue ou la décoration mal faite d'une maison. En revanche, en manipulant la sévérité des conséquences par le caractère légal vs. non légal d'un dommage, Trémolière et De Neys (2013) ont obtenu des effets de cette gravité sur le jugement moral des participants. Par conséquent, l'effet des motivations de l'agent sur l'attribution de sanctions, alors même que ces actions ne sont pas reliées au dommage, pourrait être lié à la forte volonté de blâmer les agents de nos scénarios.

Il convient cependant de souligner une limite de notre étude concernant l'opérationnalisation de l'attribution de sanctions. Nous n'avons mesuré que le degré d'accord des participants pour sanctionner l'agent dans l'absolu, et non pas pour le dommage en question. De plus, nous n'avons pas demandé d'indiquer de peine précise contrairement à Cushman (2008) ou Shultz et ses collègues (1981). Nous pourrions nous attendre à ce que les participants attribuent une amende et/ou une peine d'incarcération plus importantes lorsque l'agent est une condition nécessaire réelle au dommage. De futures recherches pourraient approfondir cette question.

Dans l'ensemble, nos travaux gardent donc la même consistance : les participants semblent respecter un minimum de critères rationnels lorsqu'il effectue des jugements de causalité et de responsabilité. Le critère de nécessité réelle est constamment respecté par nos participants à l'exception de l'attribution de sanctions générales à l'agent.

En ce qui concerne l'hypothèse d'effet de médiation du CCM (Alicke et al., 2011 ; Rogers et al., 2019), les analyses n'ont pas permis de confirmer que les réactions évaluatives vis-à-vis de l'agent médiatiseraient l'effet de ses motivations sur le jugement causal. En effet, les analyses n'ont révélé aucun effet de médiation de l'évaluation de l'agent, de ses intentions, de ses actions ou encore de la combinaison de ces trois éléments. Cependant, avant de rejeter cette hypothèse, il convient d'indiquer que des effets totaux significatifs et des effets directs et indirects non significatifs pourraient être liés à un manque de puissance statistique. Notre échantillon pour l'analyse de médiation correspondait au nombre de participants dans la condition du lien causal présent du scénario feu de forêt ($n = 40$). Or, cette taille d'échantillon ne permet de capter que les tailles d'effet larges pour des effets de médiation (voir Fritz & MacKinnon, 2007, Tableau 3). Si notre échantillon nous permettait de capter l'effet direct (Valence morale → Jugement causal), dont la taille d'effet moyenne est large (cf., analyses de

puissance a priori de l'étude 2, f de Cohen supérieur à .40), celui-ci pourrait être insuffisant pour détecter un effet de médiation petit ou moyen. Ainsi, nous ne pouvons pas conclure que des tailles d'effets petites ou moyennes n'existaient pas pour l'effet de médiation prédit par le CCM. Par ailleurs, l'absence d'effet direct peut également suggérer un problème de multicolinéarité (Zhao et al., 2010) lié à l'association très forte entre notre variable indépendante et notre variable médiatrice (cf. la très forte taille d'effet des motivations sur l'évaluation de l'agent).

Un dernier résultat notable est celui de la différence de scores d'évaluation de l'agent entre les scénarios Feu de forêt vs. Salmonelle. Cette différence pourrait notamment expliquer la non-réplication de l'effet de la valence morale dans la condition du lien causal présent du scénario de la salmonelle. En effet, si le premier scénario (i.e., feu de forêt) semble avoir montré une opérationnalisation efficace de la valence morale de l'agent, le second (i.e., salmonelle) est apparu plus problématique. Dans le scénario du feu de forêt, les données ont mis en évidence un consensus important parmi les participants pour évaluer l'agent positivement dans la condition *louable* et négativement dans la condition *blâmable*, et ce, sur chaque dimension de l'évaluation (i.e., sa personne, ses intentions et ses actions). Bien que ce consensus semble avoir été présent dans la condition blâmable du scénario de la salmonelle, aucune tendance évaluative ne pouvait être distinguée chez les participants dans la condition *louable* concernant sa personne et ses actions (cf. distributions des scores d'évaluation). Par conséquent, l'agent « louable » ne semble pas vraiment l'avoir été aux yeux de nos participants. Nous proposons deux explications non exclusives à ce phénomène. La première est que les justifications pour agir de l'agent louable seraient meilleures dans le scénario du feu de forêt que dans celui de la salmonelle. De manière générale, les justifications servent à mettre en avant le fait que le comportement de l'agent était la meilleure option parmi toutes les alternatives possibles (Greenawalt, 1986 ; Malle et al., 2014). Il devient alors plus aisé d'imaginer que, pour protéger

le village, le garde forestier était obligé d'agir face au feu menaçant. En revanche, les participants ont pu considérer que la prolifération de la salmonelle était moins probable et que l'employé avait moins de raisons d'agir pour prévenir le risque par rapport au garde forestier face au feu. Par conséquent, il serait plus facile d'excuser le garde forestier que l'employé pour ses actions. En effet, des études ont montré que, même si un agent cause un dommage, le blâme peut être atténué par des justifications de l'action effectuée (Markman & Tetlock, 2000 ; Shaver, 1985 ; Weiner, Amirkhan, Folkes, & Verette, 1987). Ceci expliquerait le consensus des participants du scénario du Feu de forêt à évaluer positivement l'agent et la grande variance dans les scores d'évaluations des participants du scénario de la salmonelle. Une seconde explication serait à nouveau la sévérité du dommage (Trémolière & De Neys, 2013). En effet, dans le scénario du feu de forêt, l'action de l'agent provoque la mort de plusieurs personnes, alors que, les clients tombent seulement malades dans le scénario de la salmonelle. Il est donc possible que le scénario de la salmonelle ne provoque pas une réaction évaluative suffisamment forte pour que l'effet de la valence morale apparaisse lorsque l'agent cause le dommage. Il pourrait ainsi être pertinent de tester ces deux mêmes scénarios en faisant varier la sévérité de chacun d'entre eux pour étudier l'apparition de l'effet de la valence morale.

Étude 4 : Effet de la valence morale sur le jugement causal en fonction du critère normatif de nécessité réelle et de la sévérité du dommage avec un agent aux motivations réellement louables

L'objectif de cette étude était donc de répliquer la précédente étude en manipulant la sévérité du dommage, et plus particulièrement son caractère létal ou non létal. Nos hypothèses restent les mêmes. Nous prédisons qu'il y aurait également une interaction entre les motivations de l'agent et la sévérité du dommage sur les jugements causaux, de responsabilité et de sanctions lorsqu'un lien causal est présent. Plus précisément, lorsqu'un lien est présent, l'effet de la valence morale apparaîtra davantage lorsque le dommage est létal plutôt que

lorsqu'il est non létal. Lorsqu'il n'y a pas de lien causal entre l'agent et le dommage, nous ne prédisons aucun effet de la valence morale et ce, quelle que soit la sévérité du dommage.

Méthode

Design. Le plan expérimental était un plan factoriel 2 (Motivations : louables, blâmables) × 2 (Lien causal : présent, absent) × 2 (Sévérité : dommage létal, dommage non létal) en inter-sujets. Les participants étaient répartis aléatoirement dans une des huit conditions de l'un des deux scénarios (Feu de forêt, Salmonelle).

Participants. Nous avons recruté 320 participants (76% de femmes, $n = 243$) par le biais du réseau social Facebook (soit 20 participants par condition). La moyenne d'âge était de 25.38 ans ($SD = 8.39$; min = 18 ; max = 56). A nouveau, aucune compensation financière n'a été versée aux participants.

Matériel et Procédure. Le matériel et la procédure étaient identiques à l'étude précédente, à l'exception de la sévérité du dommage. Chaque scénario avait en effet deux sévérités du dommage possibles : un dommage létal et un dommage non létal. Dans la condition *dommage létal*, les informations suivantes étaient précisées : « douze personnes perdirent la vie » (scénario Feu de forêt) ; « plusieurs clients tombèrent gravement malades et perdirent la vie » (scénario Salmonelle). Dans la condition *non létale*, il était précisé : « plusieurs maisons furent légèrement incendiées mais il n'y eut aucune victime » (scénario Feu de forêt) ; « plusieurs clients tombèrent légèrement malades » (scénario Salmonelle). De plus, nous n'avons proposé qu'une seule question d'évaluation morale contrairement à l'étude précédente : « Comment évaluez-vous les intentions et les actions de David envers le village ? » (de -10 « totalement blâmables » à 10 « totalement louables »).

Résultats

Évaluation de l'agent. Des tests de Mann-Whitney ont confirmé que l'agent blâmable était évalué plus négativement que l'agent louable dans chaque scénario, et ce, quelle que soit la sévérité du dommage (voir Tableau 7, ci-dessous). L'agent louable se voyait majoritairement attribuer un score positif (entre 0 et 10), alors que l'agent blâmable recevait majoritairement un score négatif (entre -10 et 0).

Tableau 7

Scores médians d'évaluation de l'agent en fonction de ses motivations et du scénario

Dommage	Feu de forêt		<i>U</i>	Salmonelle		<i>U</i>
	Louables (<i>n</i> = 40)	Blâmables (<i>n</i> = 40)		Louables (<i>n</i> = 40)	Blâmables (<i>n</i> = 40)	
Létal	5.00 [1.25, 9.42]	-8.00 [-10.00, -2.92]	163.0***	8.00 [5.75, 9.67]	-10.00 [-10.00, -6.75]	111.0***
Non létal	7.00 [5.00, 10.00]	-10.00 [-10.00, -7.75]	72.0***	5.00 [3.25, 7.00]	-8.50 [-10.00, -5.50]	85.5***

Note. Les quartiles sont entre crochets [Q1, Q3]. *** $p < .001$

Afin de tester notre hypothèse d'interaction entre l'effet de la valence morale et la sévérité du dommage lorsqu'un lien causal était présent, une ANOVA factorielle 2 (Lien causal : présent, absent) \times 2 (Motivations : louables, blâmables) \times 2 (Sévérité : dommage létal, dommage non létal) a été conduite sur les scores de jugement causal, de responsabilité et de sanctions pour chaque scénario (correction de l'alpha à .025). Les statistiques descriptives et les tests post-hoc avec correction de Bonferroni sont disponibles dans les tableaux 8 (scénario du feu de forêt) et 9 (scénario de la salmonelle). Par soucis de concision et de clarté, nous ne présenterons pas les résultats de certains effets non significatifs des ANOVA réalisées.

Jugement causal. Dans un premier temps, nous avons vérifié la réplification de l'effet du lien causal (présent, absent) sur le jugement causal pour chaque scénario. Dans les deux

scénarios, l'agent recevait effectivement un score de jugement causal plus élevé lorsqu'un lien causal était présent (scénario Feu de forêt : $M = 12.97$, $Mdn = 13.00$, $SD = 6.29$; scénario Salmonelle : $M = 15.24$, $Mdn = 15.00$, $SD = 5.17$) comparativement à la condition de lien causal absent (scénario Feu de forêt : $M = 7.03$, $Mdn = 5.00$, $SD = 5.83$; scénario Salmonelle : $M = 6.99$, $Mdn = 5.00$, $SD = 6.45$), avec respectivement $F(1, 152) = 41.44$, $p < .001$, $\eta_p^2 = .214$; $F(1, 152) = 87.05$, $p < .001$, $\eta_p^2 = .364$.

Les analyses ont également mis en évidence un effet principal des motivations de l'agent pour le scénario du feu de forêt, $F(1, 152) = 11.80$, $p < .001$, $\eta_p^2 = .072$. Comme attendu, l'agent aux motivations blâmables était davantage jugé causalement lié au dommage ($M = 15.43$, $Mdn = 17.00$, $SD = 6.37$) que l'agent aux motivations louables ($M = 10.53$, $Mdn = 10.00$, $SD = 5.21$). Les analyses statistiques n'ont en revanche pas montré d'effet significatif des motivations de l'agent sur le jugement causal dans le scénario de la Salmonelle, $F(1, 152) = 0.115$, $p = .735$, $\eta_p^2 = .001$.

Pour le scénario Salmonelle, en revanche, nous observions un effet principal de la sévérité du dommage allant dans le sens inverse de nos hypothèses, à savoir : l'agent se voyait attribuer un score plus élevé dans la condition *dommage non létal* ($M = 12.76$, $Mdn = 15.00$, $SD = 7.33$) relativement à la condition *dommage létal* ($M = 9.46$, $Mdn = 10.00$, $SD = 6.60$), $F(1, 152) = 13.93$, $p < .001$, $\eta_p^2 = .084$.

Contrairement à nos hypothèses, nous n'avons pas observé d'effet d'interaction significatif entre le Lien causal, les Motivations et la Sévérité du dommage, et ce, quel que soit le scénario. L'absence d'interaction entre le lien causal et les motivations pourrait s'expliquer par la difficulté à répliquer l'effet de la valence morale (voir Tableaux 8 et 9, p.120). Ce dernier apparaissait uniquement dans le scénario du Feu de forêt en présence d'un lien causal et d'un dommage létal. Néanmoins, contrairement à l'hypothèse spécifique de cette étude, les analyses

n'ont pas montré d'effet d'interaction entre la sévérité du dommage et les motivations de l'agent sur le jugement causal lorsqu'un lien causal était présent, $F(1, 76) = 0.70, p = .404, \eta_p^2 = .009$ (Feu de forêt) et $F(1, 76) = 0.27, p = .608, \eta_p^2 = .003$ (Salmonelle).

Jugements de responsabilité et de sanctions. Par souci de clarté, les résultats des deux scénarios seront présentés séparément pour les jugements de responsabilité et de sanctions.

Feu de forêt. Concernant les jugements de responsabilité, les analyses ont mis en évidence un effet principal du lien causal, $F(1, 152) = 25.19, p < .001, \eta_p^2 = .142$. L'agent était davantage jugé responsable lorsque son action était causalement liée à la destruction du village ($M = 11.90, Mdn = 12.0, SD = 6.11$), que lorsque son action n'était pas causalement liée au dommage ($M = 7.40, Mdn = 5.50, SD = 6.15$). Nous avons également observé un effet principal significatif des motivations sur l'attribution de responsabilité, $F(1, 152) = 21.15, p < .001, \eta_p^2 = .122$. Le pyromane souhaitant détruire le village ($M = 11.70, Mdn = 12.50, SD = 6.77$) était davantage jugé responsable que le garde forestier souhaitant le protéger ($M = 7.59, Mdn = 6.00, SD = 5.56$). Enfin, les analyses ont montré un effet principal significatif de la sévérité du dommage sur l'attribution de responsabilité, $F(1, 152) = 8.67, p = .004, \eta_p^2 = .054$. L'agent était davantage jugé responsable lorsque le dommage était non létal ($M = 10.90, Mdn = 12.00, SD = 6.82$) que lorsqu'il était létal ($M = 8.32, Mdn = 7.00, SD = 5.93$).

Comme indiqué dans le Tableau 8 (p.120), l'effet significatif des motivations sur le jugement de responsabilité a cependant été observé uniquement lorsqu'un lien causal unissait les actions de l'agent et le dommage (lien causal présent). Néanmoins, l'effet d'interaction entre le lien causal et les motivations présentait une valeur p supérieure à notre seuil alpha, $F(1, 152) = 3.64, p = .058, \eta_p^2 = .023$. Tous les autres effets d'interaction étaient bien supérieurs au seuil de significativité. A nouveau, contrairement à l'hypothèse spécifique de cette étude, les analyses n'ont pas montré d'effet d'interaction entre la sévérité du dommage et les motivations

de l'agent sur le jugement de responsabilité lorsqu'un lien causal était présent, $F(1, 76) = 0.25$, $p = .619$, $\eta_p^2 = .003$.

Concernant les sanctions, l'agent était de façon significative davantage sanctionné en présence d'un lien causal ($M = 11.00$, $Mdn = 12.00$, $SD = 7.57$), qu'en l'absence de ce lien causal ($M = 7.35$, $Mdn = 4.50$, $SD = 6.97$), $F(1, 152) = 14.47$, $p < .001$, $\eta_p^2 = .087$. De plus, l'agent blâmable ($M = 13.20$, $Mdn = 16.00$, $SD = 7.18$) était significativement plus sanctionné que l'agent louable ($M = 5.15$, $Mdn = 2.00$, $SD = 5.33$), $F(1, 152) = 70.90$, $p < .001$, $\eta_p^2 = .318$. Contrairement au jugement causal et au jugement de responsabilité, la taille d'effet des motivations sur les sanctions était bien supérieure à celle du lien causal. Les analyses statistiques n'ont en revanche pas montré d'effet significatif de la sévérité du dommage sur l'attribution de sanctions, $F(1, 152) = 1.65$, $p = .201$, $\eta_p^2 = .011$, ni d'effet d'interaction entre la sévérité et les motivations lorsqu'un lien causal était présent, $F(1, 76) = 0.55$, $p = .460$, $\eta_p^2 = .007$.

Salmonelle. L'agent était davantage jugé responsable lorsque ses actions étaient causalement liées à la propagation de la salmonelle ($M = 13.30$, $Mdn = 14.5$, $SD = 6.11$) que lorsque ce n'était pas le cas ($M = 7.11$, $Mdn = 5.00$, $SD = 6.60$), $F(1, 152) = 40.65$, $p < .001$, $\eta_p^2 = .211$. Par ailleurs, son score de responsabilité était plus élevé lorsque ses motivations étaient blâmables ($M = 11.40$, $Mdn = 13.00$, $SD = 7.57$) que lorsqu'elles étaient louables ($M = 9.00$, $Mdn = 11.00$, $SD = 6.31$), $F(1, 152) = 6.10$, $p = .015$, $\eta_p^2 = .039$. A nouveau, l'agent était davantage jugé responsable lorsque le dommage était non létal ($M = 11.30$, $Mdn = 12.50$, $SD = 7.26$) que létal ($M = 9.14$, $Mdn = 10.00$, $SD = 6.72$). En revanche, cet effet de la sévérité du dommage était supérieur à notre seuil de 2,5%, $F(1, 152) = 4.78$, $p = .030$, $\eta_p^2 = .030$. Comme pour le scénario précédent, aucun effet d'interaction n'était significatif. Contrairement à notre hypothèse, l'effet d'interaction entre la sévérité du dommage et les motivations de l'agent

n'était pas significatif dans la condition d'un lien causal présent, $F(1, 76) = 0.32, p = .576, \eta_p^2 = .004$.

Concernant l'attribution de sanctions, l'agent était davantage sanctionné en présence d'un lien causal ($M = 12.20, Mdn = 13.00, SD = 6.39$) qu'en l'absence de ce lien ($M = 8.06, Mdn = 9.00, SD = 6.71$), $F(1, 152) = 21.39, p < .001, \eta_p^2 = .123$. Il était également davantage sanctionné lorsque ses motivations étaient blâmables ($M = 13.20, Mdn = 13.00, SD = 6.57$) plutôt que louables ($M = 7.11, Mdn = 7.00, SD = 5.73$), $F(1, 152) = 45.45, p < .001, \eta_p^2 = .230$. Néanmoins, les analyses n'ont pas montré d'effet de la sévérité du dommage, $F(1, 152) = 0.38, p = .541, \eta_p^2 = .002$. Enfin, l'effet d'interaction entre le lien causal et la sévérité était statistiquement significatif, $F(1, 152) = 9.56, p = .002, \eta_p^2 = .059$.

Plus précisément, les tests post-hoc ont montré que l'agent était davantage sanctionné lorsque le dommage était non léthal ($M = 9.72, Mdn = 11.00, SD = 6.99$) plutôt que léthal ($M = 6.40, Mdn = 2.00, SD = 6.04$), dans la condition d'absence de lien causal, $t(152) = 2.62, p = .047, d = 0.59, ICd\ 95\% = [0.14, 1.03]$. Cette différence n'était pas significative dans la condition de présence d'un lien causal, $t(76) = 1.75, p = .300, d = 0.39, ICd\ 95\% = [-0.05, 0.84]$. Les autres effets d'interactions n'étaient pas significatifs. Contrairement à l'hypothèse spécifique de cette étude, l'effet d'interaction entre la sévérité du dommage et les motivations de l'agent n'était pas significatif lorsqu'un lien causal était présent, $F(1, 76) = 0.690, p = .409, \eta_p^2 = .009$.

Tableau 8

Scores moyens (écarts-types) de jugement causal, de responsabilité et de sanctions en fonction du lien causal, des motivations de l'agent et de la sévérité du dommage pour le scénario du Feu de forêt

	Présent								Absent							
	Létal				Non létal				Létal				Non létal			
	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀
Cause	9.80 (4.25)	15.80 (5.34)	1.02* [0.37, 1.67]	75.67	11.25 (6.04)	15.05 (7.39)	0.65 [0.01, 1.29]	1.06	5.40 (4.91)	6.85 (4.55)	0.25 [-0.38, 0.88]	0.45	7.20 (6.61)	8.65 (6.88)	0.25 [-0.38, 0.88]	0.37
Responsabilité	7.70 (4.80)	14.10 (5.16)	1.19** [0.53, 1.85]	104.13	10.25 (5.29)	15.45 (6.14)	0.97* [0.32, 1.62]	6.77	4.20 (3.58)	7.30 (5.41)	0.53 [-0.11, 1.16]	1.80	8.20 (6.68)	9.90 (7.23)	0.29 [-0.34, 0.92]	0.39
Sanctions	6.15 (5.58)	15.25 (6.28)	1.59*** [0.91, 2.27]	812.55	5.75 (4.80)	16.75 (6.10)	1.92*** [1.22, 2.62]	54 × 10 ³	3.80 (4.75)	9.00 (7.13)	0.82 [0.18, 1.47]	4.99	4.90 (6.17)	11.70 (6.94)	1.08** [0.42, 1.73]	15.99

Note. * p <.05 ** p <.01 *** p <.001

Tableau 9

Scores moyens de jugement causal, de responsabilité et de sanctions en fonction du lien causal, des motivations de l'agent et de la sévérité du dommage pour le scénario de la Salmonelle

	Présent								Absent							
	Létal				Non létal				Létal				Non létal			
	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀	Louables (n = 20)	Blâmables (n = 20)	d	BF ₁₀
Cause	13.10 (2.90)	14.10 (6.83)	0.20 [-0.43, 0.85]	0.36	16.80 (5.68)	16.95 (3.49)	0.03 [-0.60, 0.66]	0.31	4.35 (4.38)	6.30 (5.68)	0.32 [-0.31, 0.95]	0.55	10.65 (6.91)	6.65 (7.18)	0.65 [0.01, 1.29]	1.08
Responsabilité	10.80 (5.20)	14.80 (6.78)	0.67 [0.03, 1.31]	1.68	12.50 (4.72)	15.00 (6.87)	0.42 [-0.21, 1.05]	0.63	3.10 (3.65)	7.85 (4.97)	0.76 [0.12, 1.40]	23.57	9.60 (7.06)	7.90 (8.28)	0.27 [-0.90, 0.36]	0.38
Sanctions	10.70 (4.46)	15.95 (6.26)	0.95** [0.30, 1.60]	9.93	7.45 (5.13)	14.75 (6.02)	1.32*** [0.66, 1.99]	122.27	2.55 (3.39)	10.25 (5.67)	1.32*** [0.65, 1.99]	2225.56	7.75 (6.53)	11.70 (7.04)	0.68 [0.04, 1.32]	1.15

Note. * p <.05 ** p <.01 *** p <.001

Discussion de l'Étude 4

Cette étude a, une nouvelle fois, confirmé l'importance du critère de nécessité réelle pour le jugement causal de sens commun, mais également pour le jugement de responsabilité. Ces résultats signifient que les individus respecteraient la présence d'un lien causal réel entre un agent et un dommage pour le désigner comme cause et responsable, et ce, même lorsque cet agent est évalué très négativement et qu'un dommage légal a eu lieu. Notons par ailleurs que le lien causal unissant l'agent au dommage présentait une forte taille d'effet sur ces deux jugements (causalité, responsabilité), et ce, quel que soit le scénario utilisé (Feu de forêt, Salmonelle) ou la sévérité du dommage à évaluer (légal, non légal). Concernant les prédictions du CCM, l'effet des motivations de l'agent sur le jugement causal a été retrouvé uniquement dans le scénario du Feu de forêt, et ce, lorsqu'un lien causal unissait l'agent et le dommage et que le dommage était légal. Il s'agissait ainsi d'une condition dans laquelle le participant pouvait, face à un événement suffisamment grave (i.e., la mort de plusieurs habitants), formuler un jugement plus sévère à l'endroit de l'agent blâmable tout en respectant le critère de nécessité réelle. L'effet de la valence morale n'a été répliqué ni dans les autres conditions de ce scénario ni dans celles du scénario de la Salmonelle. Malgré une opérationnalisation propice à son apparition (i.e., dommage légal et contraste important entre un agent réellement louable et un agent réellement blâmable), ses conditions d'apparitions semblent limitées à un minimum de rationalité et à une évaluation négative suffisamment importante vis-à-vis de l'événement.

Un second résultat notable est celui de l'impact de la sévérité du dommage sur les jugements causaux dans le scénario de la Salmonelle. L'effet était inverse de ce que prédit le CCM. En effet, selon ce modèle, la sévérité du dommage devrait augmenter les évaluations spontanées négatives vis-à-vis de l'agent blâmable et ainsi favoriser le *processus de validation du blâme*. Or, ce que nous observons est que, l'employé du restaurant était davantage jugé

comme cause de la propagation de la Salmonelle lorsque les clients tombaient légèrement malade (dommage non légal), relativement à la situation où les clients décédaient à la suite de l'intoxication alimentaire (dommage légal). Nous voyons deux explications à ce résultat : les justifications à agir (que nous avons déjà évoquées dans la discussion de l'Étude 3) et la prévisibilité d'un dommage légal.

En effet, la prévisibilité du dommage par l'agent peut également avoir eu un effet sur le jugement causal. Les résultats de Fincham et Jaspars (1983), comme ceux de Lagnado et Channon (2008) ont montré que lorsqu'un dommage est imprévisible (vs. prévisible), l'agent est jugé moins causal et moins responsable de celui-ci. Dans notre étude, il est probable que les participants aient jugé les décès faisant suite à un feu criminel comme plus probables, et donc plus prévisibles, que des décès provoqués par le simple réchauffement d'une chambre froide. Dans cette perspective, on peut s'attendre à ce que l'employé du restaurant prévoyait de causer des intoxications alimentaires légères plutôt que létales. Par conséquent, l'employé décrit dans la condition *dommage non légal* peut avoir été jugé plus causal et responsable que celui décrit dans la condition *dommage légal*, car le dommage non légal était plus prévisible que le dommage légal.

Les résultats concernant les jugements de responsabilité et de sanctions semblent aller dans le sens de cette interprétation. L'agent est davantage désigné comme responsable et sanctionné lorsque le dommage est non légal (vs. légal) dans les deux scénarios. À l'inverse d'un raisonnement motivé, nous aurions ainsi des jugements respectueux, à la fois du critère de nécessité réelle pour la responsabilité, mais également de certaines normes juridiques telles

que la prise en compte de la prévisibilité du dommage³⁹ ou des justifications à agir⁴⁰. Ainsi, même lorsque les motivations à agir de l'agent et la sévérité du dommage favorisaient des évaluations négatives vis-à-vis de l'agent, les jugements de causalité, de responsabilité et de sanctions semblent se conformer aux critères normatifs de la causalité et/ou du droit.

En somme, cette difficulté à répliquer l'effet de la valence morale ainsi que les résultats sur la sévérité du dommage apportent relativement peu de soutien à l'hypothèse d'un jugement causal biaisé par une évaluation négative. En effet, cette étude corrobore les résultats obtenus dans les études précédentes, à savoir que le sens commun semble s'écarter relativement peu des critères normatifs du jugement causal. Et ce, malgré une opérationnalisation propice à l'apparition de l'effet de la valence morale. En revanche, les prédictions du CCM ne sont pas totalement écartées car l'effet de la valence morale apparaît dans certaines conditions. Par conséquent, les jugements de sens commun ne peuvent pas être considérés comme parfaitement rationnels.

³⁹ Par exemple, l'article 1231-3 du code civil prévoit que la responsabilité contractuelle n'est engagée que si le dommage était prévisible.

⁴⁰ L'article 122-7 du code pénal précise la justification des comportements et notamment la nécessité d'agir proportionnellement face à une menace plus ou moins importante.

Discussion du Chapitre 3

L'objectif de ce chapitre était d'étudier dans quelle mesure les jugements causaux de sens commun pouvaient s'écarter des critères normatifs de la causalité (i.e., nécessité réelle et sensibilité), lorsqu'un événement impliquant un agent humain suscite à son égard une évaluation spontanée négative. En d'autres termes, jusqu'où le *processus de validation du blâme*, tel que théorisé par le CCM, pourrait biaiser l'évaluation des liens causaux ? La réponse à cette question permettrait de placer le jugement causal de sens commun sur le continuum de la rationalité dans ce type de situations et de délimiter les conditions d'apparition de l'effet de la valence morale. Une synthèse des principaux résultats des quatre études conduites pour répondre à cette question est présentée dans le Tableau 10 ci-dessous.

Tableau 10

Synthèse de la réplication de l'effet de la valence morale sur le jugement causal dans les études 1, 2, 3 et 4

Étude 1			Étude 2			Étude 3		Étude 4				
Lien fort	Lien faible	Lien absent	Lien fort	Lien faible	Lien absent	Lien présent	Lien absent	Lien présent		Lien absent		
								Létal	Non létal	Létal	Non létal	
						Scénario						
X	-	-	X	X	-	Feu de forêt	X	-	X	-	-	-
						Salmonelle	-	-	-	-	-	-

Un premier enseignement important de nos études concerne l'effet des modèles causaux sur les jugements de causalité. Les résultats de nos quatre études montrent en effet que les principaux déterminants du jugement causal de sens commun sont les deux critères normatifs de la causalité, à savoir la nécessité réelle (Hilton et al., 2021 ; Icard et al., 2017) et la sensibilité (Icard et al., 2017 ; Lewis, 1973 ; Woodward, 2006, 2010). Plus précisément, nous avons observé que l'agent était considéré comme la cause du résultat lorsque son action était une condition nécessaire réelle au dommage (liens causaux fort, faible vs. lien causal absent) (Études 1 à 4). De plus, l'agent était plus susceptible d'être identifié comme la cause réelle, et

de se voir attribuer un score de causalité plus élevé, lorsque son action était insensible aux changements des circonstances de fond (lien causal fort) plutôt que lorsqu'elle était sensible à de tels changements (lien causal faible) (Études 1 et 2).

Le second enseignement porte sur l'interaction entre le modèle causal et la valence morale dans les jugements causaux. Nos résultats nous amènent à considérer que, même en présence d'une forte propension à blâmer l'agent, les jugements causaux s'écartent très peu des critères normatifs de la causalité. Ainsi, des déviations à la rationalité, caractérisées par l'effet de la valence morale, ont bien été observées dans des conditions parfaitement circonscrites par le critère de nécessité réelle. En somme, il semblerait que les individus n'iraient pas jusqu'à inventer un lien causal inexistant entre un agent et un dommage, et ce, même si l'agent souhaitait provoquer le dommage, a agi pour provoquer le dommage, et que le dommage souhaité s'est produit.

Les profanes semblent par ailleurs prendre en considération le critère normatif de force causale (i.e., sensibilité aux circonstances) et ce, même lorsqu'un agent a des motivations blâmables (e.g., tuer son mari pour des raisons financières) (Études 1 et 2). Les résultats de l'Étude 1 ont notamment montré que, lors d'une tâche de jugement causal mobilisant une question ouverte, nous ne répliquions pas l'effet de la valence morale lorsque l'agent était sensible aux circonstances (i.e. faible lien causal). En effet, dans cette condition, les participants désignaient davantage la condition ayant la plus grande force causale (i.e., l'allergie provoquée par le nouveau plat vs. l'agent voulant tuer pour des raisons financières). Il semblerait ainsi que le critère de sensibilité, comme celui de nécessité réelle, soit davantage déterminant pour les jugements causaux que l'évaluation négative suscitée par l'agent.

Le fait que nous ayons retrouvé l'effet de la valence morale dans la condition du lien causal faible de l'Étude 2, avec une question à échelle centrée sur l'agent, nous conduit

également à questionner l'influence du format de la tâche de jugement causal, et plus précisément la nature de la question de causalité. En effet, lorsque les participants devaient désigner la cause de l'événement (question ouverte, Étude 1), l'effet de la valence morale n'était pas reproduit lorsque le lien causal entre l'action de l'agent et le dommage était faible. Cependant, lorsque les participants ont été invités à donner une estimation plus fine de la contribution causale de l'agent au dommage (question fermée centrée sur l'agent, Études 2 à 4), nous avons reproduit l'effet de la valence morale lorsque le lien causal entre l'action et le dommage était faible⁴¹. Ce dernier résultat est davantage conforme aux prédictions du CCM et aux études qui ont testé l'effet de la valence morale avec un lien causal faible (Alicke, 1992 ; Rogers et al., 2019). En effet, selon le CCM, on devrait observer cet effet lorsque le lien causal entre l'agent et le dommage est ambigu (*lien faible*), et non lorsque l'agent est évidemment un contributeur au dommage (*lien fort*). Toutefois, même si l'effet de la valence morale est apparu dans le *lien faible* de l'Étude 2, les scores étaient globalement plus bas que dans la condition du *lien fort*. Cela indique, que les participants respecteraient le critère de force causale même dans des conditions propices au blâme (voir infra pour une discussion sur cette inconsistance avec le CCM).

Ces résultats apportent un éclairage majeur tant sur le plan théorique qu'appliqué. D'un point de vue théorique, il apparaît que les « biais » observés dans les tâches qui nous intéressent ne sont pas dus à des problèmes de connaissances a priori (comme c'est le cas pour certains biais, voir Stanovich, 2018). De façon conforme à de précédents résultats sur le jugement causal (Grinfeld et al., 2020 ; Hilton et al., 2021 ; Vasilyeva et al., 2018), nos résultats montrent que les individus connaissent et mobilisent les critères normatifs dans des situations à l'issue négative impliquant un agent humain. Par ailleurs, nos études ont montré que les profanes ne

⁴¹ Ceci vaut uniquement pour le scénario Feu de forêt dans les études 3 et 4.

se contentent pas seulement de connaître et de mobiliser les deux critères normatifs, ils les appliquent de manière prioritaire. En effet, comme nous l'avons évoqué, il y aurait une primauté des critères normatifs de la causalité sur l'influence des évaluations négatives en ce qui concerne les jugements causaux de sens commun, et ce, même dans des conditions très favorables à l'apparition de l'effet de la valence morale (contraste plus important entre la valence morale des motivations via une opérationnalisation plus favorable des motivations louables et dommage légal). Ces résultats tendent ainsi davantage vers la corroboration d'une conception « optimiste » de la rationalité humaine en ce qui concerne les jugements causaux de sens commun.

D'un point de vue appliqué, cela indique qu'une formation à ces critères normatifs ou à leur application ne semble pas indiquée. Ce constat nous laisse à considérer deux autres options pour limiter l'effet de la valence morale sur le jugement causal. Ces deux options renvoient notamment à la distinction de Fischhoff (1982) entre les problèmes de juges (le biais vient de juges défaillants) et les problèmes de tâche (le biais vient d'une tâche défaillante). En d'autres termes, soit il s'agit bien d'un biais cognitif qui apparaît dans certaines circonstances (i.e., juges défaillants), soit le biais est inhérent à la tâche (e.g., l'interprétation que le participant se fait de la tâche n'est pas conforme aux attentes de l'expérimentateur).

Le CCM s'inscrit dans la première approche : ce sont les juges qui sont défaillants et biaisés par leurs évaluations spontanées négatives. Selon cette perspective, il faut donc agir sur les participants pour neutraliser l'effet de la valence morale sur le jugement causal. Alicke considère que les évaluations spontanées sont automatiques mais peuvent être annulées par un effort conscient (2000, p.568). Cet effort conscient renvoie aux travaux sur l'inhibition des processus non-pertinents à la réalisation de la tâche (Houdé, 2019 ; Houdé et al., 2000 ; Moutier & Houdé, 2003). Ce problème d'inhibition d'un processus parasite peut avoir différentes

origines telle qu'un manque de motivation ou de ressources cognitives pour mettre en place un processus d'inhibition coûteux, ou encore une incapacité à détecter l'incompatibilité du processus avec la tâche (pour une revue, De Neys & Bonnefon, 2013). Si l'origine du biais est celle-ci, il conviendrait d'inciter les individus à inhiber leurs évaluations négatives lorsqu'ils effectuent un jugement causal. Une fois ces évaluations inhibées, les jugements causaux devraient alors se conformer exclusivement aux critères normatifs, puisque nos résultats semblent appuyer l'idée qu'ils sont maîtrisés par les individus. Si cette inhibition est possible et appliquée, nous pourrions espérer obtenir des jugements causaux parfaitement rationnels de la part des profanes.

Néanmoins, la différence de résultats observée selon la nature de la question causale nous incite à davantage considérer un problème de tâche qu'un problème de juges. Pour rappel, la comparaison des études 1 et 2 indique que l'effet de la valence morale est davantage retrouvé avec une question fermée centrée sur l'agent qu'avec une question ouverte. Comme précédemment présenté (cf. pp.85-86, discussion de l'Étude 1), ce résultat pourrait s'expliquer par le fait qu'une question à échelle (Étude 2) est plus précise pour capter un effet par rapport à une question ouverte impliquant une réponse dichotomique (Étude 1). Cependant, la comparaison des jugements où l'agent est sélectionné comme la cause du dommage (i.e., mention de l'agent en réponse à la question ouverte, score de causalité en réponse à la question fermée avec échelle de réponse Likert) montre que lorsque la question est centrée sur l'agent (Étude 2), ce dernier est plus fortement avancé comme la cause, et ce, quel que soit la force du lien causal. Prenons les données de l'Étude 1 pour illustration : à une tâche de jugement où la question n'est pas centrée sur l'agent, 19.7% des réponses des participants ciblaient l'agent blâmable comme cause principale du dommage dans la condition d'un lien fort. Or, lorsque la tâche de jugement mobilisait une question centrée sur l'agent (Étude 2), nous observions que la médiane des scores de jugement causal était de 21 pour l'agent blâmable dans cette même

condition, soit la réponse maximale de causalité sur l'échelle proposée (1 « pas du tout la cause » à 21 « totalement la cause »). Ces résultats semblent suggérer que les participants considèrent massivement l'agent comme cause principale lorsqu'une question fermée centrée sur l'agent est utilisée, mais peu lorsqu'il s'agit d'une question ouverte.

Le prochain chapitre se concentrera sur cet effet des questions sur la sélection causale. En effet, comme d'autres auteurs (e.g., Driver, 2008 ; Samland & Waldmann, 2016), nous nous appuyerons sur les travaux relatifs à la perspective pragmatique du jugement causal en considérant que ceux-ci ne sont pas liés à des juges défaillants mais à la nature de la tâche de jugement de causalité qui pourrait, comme défendu par certains, être une tâche ambiguë. Selon cette perspective, la réduction de l'effet de la valence morale pour tendre vers des jugements causaux rationnels passerait, comme nous le verrons, par une clarification de la tâche de jugement causal.

Chapitre 4. Une perspective pragmatique de l'effet de la valence morale

*« In fact, I shall argue that when the correct normative standards are applied, human causal inference appears to be more rational than was once thought. »
Hilton (2007, p.5)*

L'approche que nous avons adoptée jusqu'à présent consistait à comparer le jugement effectif des individus à des critères normatifs logiques, philosophiques et juridiques. Néanmoins, plusieurs auteurs ont critiqué cette méthode pour évaluer la rationalité des jugements de sens commun (pour une revue, Wagner-Egger, 2011). Pour une partie d'entre eux (e.g., Hilton, 1991, 1995 ; Hilton & Erb, 1996 ; Mercier & Sperber, 2011 ; Van der Henst, 2002), les chercheurs utilisant cette approche commettent l'erreur de considérer que la bonne résolution d'une tâche de raisonnement est une activité essentiellement individuelle consistant à appliquer un modèle normatif (i.e., logique, mathématique, etc.) sur les informations littérales de la tâche. Or, la plupart des informations sur lesquelles nous raisonnons nous parviennent par un média social (e.g., expérimentateur, juge, journaliste, ami). Ce média social possède ses propres objectifs et souhaite acquérir des informations bien spécifiques. Par conséquent, s'ajoute à la réalisation de la tâche de raisonnement, un processus de décryptage du contenu du message et d'inférences des intentions de notre interlocuteur, notamment celles de l'expérimentateur (Hilton et al., 2017). Selon cette perspective, le modèle normatif auquel le chercheur compare le jugement de sens commun doit donc intégrer, en plus de la logique classique, la logique de la conversation (Austin, 1962 ; Grice, 1975 ; Levinson, 1983 ; Sperber & Wilson, 1996). La logique de la conversation possède ses propres règles telles que les maximes de Grice (1975) : la maxime de quantité (i.e., ne pas fournir trop peu ou trop d'informations), la maxime de relation (i.e., être pertinent par rapport à la question posée), la

maxime de modalité (i.e., être clair) et la maxime de qualité (i.e., affirmer ce que l'on juge vrai). Cette perspective d'intégrer la logique de la conversation au raisonnement de sens commun peut être qualifiée d'approche pragmatique du raisonnement. La pragmatique se définit ici comme l'étude cognitive, sociale et culturelle du langage et de la communication (Bernicot, 2002). Au regard de nos précédents résultats et de travaux récents que nous détaillerons par la suite, le présent chapitre adoptera cette perspective vis-à-vis de l'effet de la valence morale.

Dans un premier temps, nous décrirons brièvement les règles conversationnelles et dans quelle mesure le fait de ne pas les prendre en compte peut amener, à tort, à la conclusion d'un biais de raisonnement chez les participants. Dans un second temps, nous verrons comment l'approche pragmatique a été appliquée au domaine du jugement causal en s'intéressant à l'ambiguïté du mot *cause*. Nous verrons également que certains auteurs ont démontré l'influence du type de mesure (i.e., questions posées) sur l'apparition de l'effet de la valence morale. Nous élargirons cette analyse à travers une revue narrative de la littérature en nous intéressant à la structure des scénarios utilisés dans les recherches conduites à ce jour sur l'effet de la valence morale. Plus précisément, nous verrons comment le modèle causal des scénarios en interaction avec la question posée (mentionnant, ne mentionnant pas l'agent) pourrait avoir des implications pragmatiques sur l'apparition de l'effet de la valence morale. Enfin, nous présenterons deux études expérimentales visant à étudier l'effet du cadrage conversationnel sur l'apparition de l'effet de la valence morale.

La logique conversationnelle et les biais de raisonnement

L'approche pragmatique du raisonnement part notamment du postulat que lorsque des locuteurs communiquent, ils respectent un principe de coopération vis-à-vis de l'interaction qui a lieu (Grice, 1975). Pour ce faire, ils tâchent d'être les plus pertinents possible et partent d'une

présomption de pertinence de la part de leur interlocuteur (Levinson, 1983 ; Sperber & Wilson, 1996). Cette volonté de fournir des énoncés pertinents et la croyance que l'autre possède cette même volonté permet aux individus de s'engager dans un processus de décryptage des énoncés de l'autre (Hilton et al., 2017). Ces inférences conversationnelles, visant à formuler des hypothèses sur les intentions et le sens des mots de notre interlocuteur, constituent un raisonnement en soi. Par conséquent, une tâche de raisonnement implique, en plus de l'application des critères normatifs classiquement employés, le respect des règles conversationnelles, telles que fournir des informations nouvelles et pertinentes à l'interlocuteur (Grice, 1975 ; Sperber & Wilson, 1996). Pour illustrer la distinction entre l'application stricte et exclusive des critères normatifs de la logique classique et le respect des règles conversationnelles, imaginons l'échange suivant (proposé par Hilton, 1995) :

« - *Comment va votre femme ?*

- *Assez mal, je le crains.*

- *Et comment va la famille ?*

- *Très bien, merci ! »*

Du point de vue de la logique classique, « femme » appartient à l'ensemble « famille ». Il paraît donc irrationnel au premier abord de fournir une réponse totalement différente aux deux questions : si toute la famille de notre interlocuteur va bien, alors sa femme aussi. Si nous comparons l'échange ci-dessus avec un modèle normatif de logique formelle, la réponse du répondant sera qualifié de biaisé. Néanmoins, si nous considérons la logique de la conversation, le principe de pertinence (ou maxime de relation) implique de fournir des informations qui ne sont pas déjà connues par notre interlocuteur. Par conséquent, le mot « famille » est réinterprété dans le contexte de la conversation en soustrayant « femme » à « famille » qui ne contient désormais plus que les enfants par exemple. Le répondant a, ainsi,

modifié de manière coopérative et rationnelle (i.e., par rapport aux règles de la conversation) ses réponses en fonction de l'interaction (Hilton, 1995).

Comme dans cet exemple, plusieurs travaux ont démontré que certains « biais » étaient neutralisés lorsque les facteurs conversationnels étaient contrôlés (e.g., Schwarz, 1995 ; Schwarz et al., 1991 ; Tetlock et al., 1996). Par exemple, dans le domaine de l'attribution causale, Slugoski et ses collègues (1993) ont demandé à des participants de lire l'histoire détaillée d'un jeune qui avait sévèrement blessé l'un de ses camarades lors d'une bagarre. Le cas comprenait des informations sur la personnalité du jeune (e.g., arrogant), et des informations situationnelles sur les circonstances dans lesquelles l'incident avait eu lieu (e.g., sa mère est malade). Ils demandèrent ensuite aux participants d'expliquer l'incident à un interlocuteur qui, soit connaissait beaucoup de choses sur la personnalité du jeune mais pas les circonstances entourant le comportement violent en question, soit connaissait peu de choses de la personnalité du jeune mais connaissait bien les circonstances de son comportement. Les résultats ont montré que les participants changeaient leurs explications de façon à compléter le point de vue de leur interlocuteur. Plus précisément, ils fournissaient des explications qui insistaient sur les facteurs situationnels aux interlocuteurs qui disposaient déjà de l'information sur la personnalité du garçon, et des explications personnologiques aux interlocuteurs qui connaissaient les facteurs situationnels⁴². Ainsi, les participants ne se contentaient pas uniquement de trouver la réponse au problème qu'on leur posait, ils essayaient de restituer les informations les plus pertinentes pour l'expérimentateur. Autrement dit, la vérité n'est pas le seul objectif des raisonneurs. Ils doivent également apporter une contribution utile à la conversation. Ces résultats montreraient que l'erreur fondamentale d'attribution⁴³ pourrait

⁴² Dans cette étude, les analyses ont été effectuées sur le score moyen de deux juges indépendants en aveugle qui évaluaient les explications des participants sur une échelle de -3 (attribution purement dispositionnelle), à 0 (explication conjonctive), à 3 (attribution purement situationnelle).

⁴³ Tendance générale à surestimer l'importance des facteurs dispositionnels au détriment des facteurs situationnels (Ross, 1977, p.184).

parfois être due à une volonté de communiquer des informations nouvelles à l'interlocuteur. Ceci vient corroborer l'approche pragmatique selon laquelle au moins une partie des « biais » identifiés dans la littérature pourrait être liée à la non-prise en compte de la dimension pragmatique par les chercheurs, et notamment des critères normatifs de la conversation (i.e., qualité, quantité, pertinence, clarté), plutôt que la manifestation d'un défaut du raisonnement humain.

L'affirmation selon laquelle les chercheurs doivent tenir compte de l'interprétation des participants vis-à-vis des informations qui leur sont fournies semble aujourd'hui généralement acceptée. Le débat s'est déplacé sur la place de l'inférence conversationnelle rationnelle (vs. erreurs de jugement) dans les explications des jugements « irrationnels » relevés dans la littérature (Hilton et al., 2017). Comme abordé dans la discussion du chapitre 3, l'une et l'autre de ces explications ont des applications pratiques différentes. Si l'effet observé est lié à un biais de jugement, il convient de mettre au point des méthodes de débiaisement (e.g., Morewedge et al., 2015 ; Sellier et al., 2019). A l'inverse, si l'effet observé est lié à une inférence conversationnelle rationnelle de la part des participants, il est nécessaire d'améliorer la communication et la présentation des informations dans la tâche de raisonnement (e.g., Hilton et al., 2017 ; Sedlmeier & Hilton, 2012). Dans ce qui suit, nous verrons que la question s'est posée tant pour le jugement causal que pour l'effet de la valence morale.

L'approche pragmatique du jugement causal

Les tenants de l'approche pragmatique du jugement causal (Hilton, 1990, 1991 ; Hilton & Erb, 1996 ; Samland & Waldmann, 2016) suggèrent de distinguer deux dimensions dans le jugement causal : (1) la représentation mentale du modèle causal de l'événement et (2) la sélection de la condition la plus pertinente à mentionner comme cause du dommage dans une conversation. Chaque dimension devra être évaluée avec des critères normatifs qui lui sont

propres. Pour évaluer la rationalité de la représentation du modèle causal de l'événement, c'est le critère de « vérité » qui compte, c'est-à-dire le degré par lequel la représentation des individus est proche de la « réalité ». Nous pouvons ainsi considérer que les critères normatifs correspondants sont la nécessité réelle et la sensibilité des conditions. En ce qui concerne la dimension de sélection causale, elle ne peut pas être évaluée uniquement à travers le critère de « vérité ». Comme nous l'avons vu avec l'étude de Slugoski et ses collègues (1993), celle-ci est également concernée par une dimension pragmatique. Les explications causales ne doivent pas seulement être vraies, elles doivent être informatives et pertinentes pour l'individu qui pose la question (Hilton & Erb, 1996). En d'autres termes, les individus doivent sélectionner, au sein de leur représentation du modèle causal de l'événement, le facteur le plus pertinent à mentionner comme cause dans une conversation.

Lors de cette seconde étape d'explication causale, la difficulté pour le répondant est que le mot *cause* est un mot polysémique (Samland & Waldmann, 2016). Cette polysémie est liée à la double fonction du jugement causal : expliquer un événement ou désigner un responsable à blâmer et à sanctionner. Ces deux fonctions ont été théorisées respectivement sous le nom d'*enquête explicative* et d'*enquête attributive* (Hart & Honoré, 1985). Par exemple, la question « Quelle est la cause de l'explosion de la navette spatiale Challenger ? » peut engendrer au moins deux inférences chez l'individu à qui elle est posée. Cet individu peut être tenté de répondre « l'effet du froid sur les joints toriques des propulseurs » ou bien « la négligence de la NASA et de l'entreprise de construction des joints ». La question peut ainsi être interprétée comme une demande d'explication d'un lien causal au sens faible, à savoir « Quel est le mécanisme physique qui a littéralement fait exploser la navette ? ». Ici, l'individu pourrait répondre « l'effet du froid sur les joints toriques des propulseurs ». Cette réponse serait pertinente pour expliquer l'événement à quelqu'un qui n'a aucune information (i.e., enquête explicative). Cependant, la question peut aussi être interprétée comme une demande de

désignation d'un ou des responsables vis-à-vis de l'accident (i.e., enquête attributive), auquel cas l'individu pourrait répondre « la négligence de la NASA et de l'entreprise de construction des joints ». Le répondant indique ici que la NASA ou l'entreprise doivent être désignés comme responsables de l'accident et éventuellement sanctionnés. Le choix entre l'une des deux interprétations dépendra de facteurs contextuels pragmatiques, c'est-à-dire ce qui fait le plus sens dans la situation.

Dans cette perspective, des auteurs ont proposé que l'effet de la valence morale serait la manifestation d'une interprétation de la tâche comme une enquête attributive, c'est-à-dire de recherche d'un responsable à blâmer et/ou à sanctionner (Driver, 2008 ; Samland & Waldmann, 2016). Comme nous l'avons vu dans le Chapitre 1, l'attribution d'une responsabilité nécessite un lien causal entre l'agent et le dommage mais également, contrairement à l'enquête explicative, la prise en compte des états mentaux de l'agent tels que ses intentions ou ses motivations (Bordel et al., 2006 ; Gailey & Falk, 2008 ; Malle et al., 2014). Par conséquent, l'impact de la valence morale des états mentaux de l'agent sur le jugement causal des individus serait lié à l'interprétation du mot *cause* en termes de responsabilité (i.e., *enquête attributive*) et non en termes de causalité au sens strict (i.e., *enquête explicative*). Contrairement à l'hypothèse du CCM, l'effet de la valence morale ne serait pas la manifestation d'une représentation biaisée du modèle causal de l'événement, mais pourrait traduire une interprétation de la question comme une recherche de responsable (comprenant à la fois les dimensions objectives et subjectives de l'acte). Par conséquent, la méthode la plus pertinente à utiliser pour réduire l'effet de la valence morale serait moins d'inciter les individus à inhiber une évaluation négative à l'endroit de l'agent, que de clarifier le cadre conversationnel en précisant que la tâche est une *enquête explicative* et non une *enquête attributive*.

Cette hypothèse a été corroborée par une série de travaux menée par Samland et Waldmann (2014, 2015, 2016). Ces chercheurs ont démontré que l'effet de la valence morale n'était pas répliqué lorsque des informations incitant les participants à interpréter le mot *cause* en termes de causalité au sens strict (i.e., *enquête explicative*) étaient fournies. Une de leur hypothèse est que les questions qui se réfèrent à un agent (e.g., Dans quelle mesure l'agent a-t-il causé le dommage ?) favorisent l'interprétation du mot *cause* en termes de responsabilité. En revanche, les questions se référant à des objets (e.g., Dans quelle mesure la balle a-t-elle causé le dommage) favorisent l'interprétation du mot *cause* en termes de causalité au sens strict. Dans leur Étude 2, Samland et Waldmann (2015) ont par exemple demandé aux participants d'indiquer la cause d'un dommage en faisant varier le libellé des modalités de réponses. Ils ont demandé aux participants de lire un scénario adapté de la *pen vignette*. Il s'agissait d'un récit dans lequel un système de livraison automatique de stylos a été mis en place dans deux départements de philosophie d'une Université. Sur chaque bureau, il y a trois boutons de couleurs différentes qui permettent de commander des fournitures de bureau comme des stylos ou des gommes. Les participants lisaient ensuite, qu'en raison d'une décision budgétaire, seuls les employés du département B étaient autorisés à utiliser le système de livraison pour commander des fournitures de bureau. En revanche, les employés du département A n'étaient plus autorisés à appuyer sur les boutons situés sur leur bureau. Malgré ces nouvelles règles, les deux départements continuaient à commander des fournitures de bureau, ce qui était fréquemment critiqué par la réceptionniste. Enfin, les participants apprenaient que :

« Un matin, par hasard, Mme Smith du département A, Mme Cooper du département B et M. Wall du département B appuient sur un bouton dans leur bureau exactement à la même heure [9 h 26] :

- Mme Smith appuie sur le bouton vert et un stylo est livré dans son bureau.
- Mme Cooper appuie sur son bouton bleu et un stylo est livré dans son bureau.
- M. Wall appuie sur son bouton jaune et une gomme est livrée à son bureau.

Quelques minutes plus tard [9 h 31], la réceptionniste a besoin d'un stylo et appuie sur son bouton stylo... mais il n'y a plus de stylos en réserve. »

Les participants devaient ensuite indiquer la cause de l'absence de stylos. Les réponses possibles de la première condition étaient des agents humains (Mme Smith, Mme Cooper ou M. Wall) qui pouvaient donc se voir attribuer une responsabilité et des sanctions. Dans la seconde condition, il s'agissait des objets physiques manipulés par ces mêmes agents humains (bouton vert, bouton bleu ou bouton jaune), renvoyant ainsi davantage à la causalité au sens strict. Les résultats ont montré que l'agent humain qui enfreignait une norme prescrite (i.e., Mme Smith qui a pressé le bouton vert) était davantage désigné comme cause du dommage uniquement dans la première condition qui mentionnait des agents humains. Il n'y avait pas d'effet de l'infraction d'une norme sur le jugement causal lorsque la question se référait aux objets manipulés par les agents. Ces résultats corroboraient l'hypothèse selon laquelle l'effet des normes serait la manifestation d'une interprétation de la question en termes de responsabilité, plutôt que d'une altération de la représentation du lien causal entre l'agent et le dommage.

A l'instar des travaux de Samland et Waldmann, les résultats de nos études 1 et 2 semblent confirmer l'importance du type de question posée. En effet, nous avons vu que l'effet de la valence morale se manifestait davantage avec une question mentionnant l'agent (e.g., Dans quelle mesure l'agent a-t-il causé le dommage ? Étude 2) qu'avec une question ne mentionnant pas l'agent (e.g., Quelle est la cause principale du dommage ? Étude 1). Dans la prochaine section, nous proposerons d'étendre cette réflexion pragmatique à propos du type de mesure du jugement causal en prenant également en compte les scénarios présentés aux participants, notamment leur modèle causal. En effet, nous proposerons que le type de lien causal entre l'agent et le dommage pourrait également avoir un impact sur les inférences conversationnelles des participants.

Les implications pragmatiques des scénarios et de la focalisation de la question

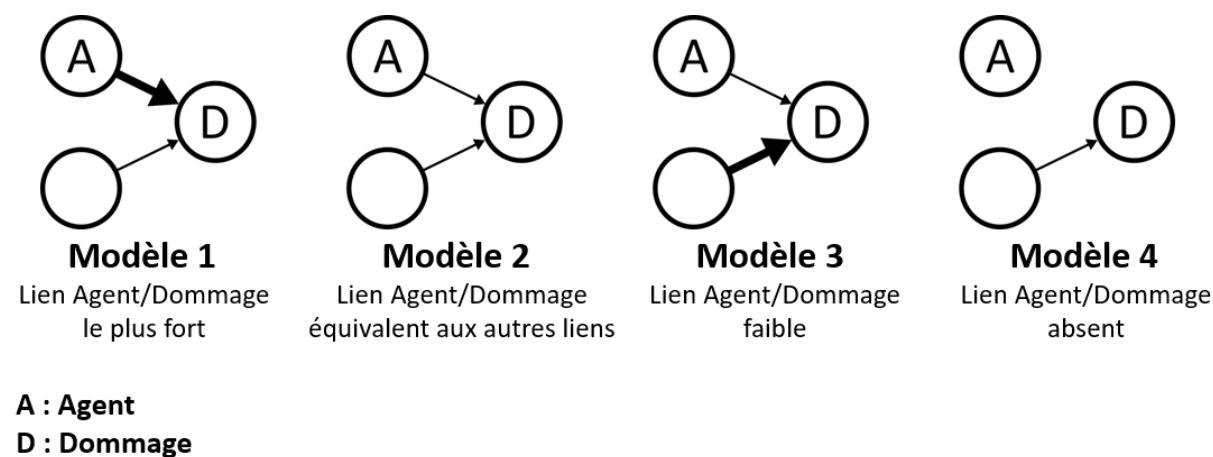
Lors d'une tâche de jugement causal, les participants se trouvent face à un ensemble d'informations fournies par l'expérimentateur, généralement sous la forme d'un texte, vis-à-vis desquelles ils vont devoir se représenter le modèle causal de la situation (Graesser et al., 1994 ; Schank & Abelson, 1988 ; Trabasso & Bartolone, 2003 ; Trabasso & Sperry, 1985 ; Trabasso & van den Broek, 1985). Comme nous l'avons vu dans le chapitre 1, un modèle causal comprend la structure causale (i.e., présence ou absence d'un lien causal) et la force causale (i.e., force de ce lien). Ainsi, chaque situation va posséder un modèle causal singulier. Il en est naturellement de même dans un contexte judiciaire, où chaque affaire a son propre modèle causal : l'accusé peut avoir causé ou non le dommage, ce qui définit la structure causale et, si cette relation causale existe, le lien entre ses actions et le dommage peut être plus ou moins important, ce qui définit la force causale. Néanmoins, un modèle causal laissera toujours de côté de nombreux détails (Lagnado & Gerstenberg, 2017). D'une part, parce que nous n'avons pas les capacités cognitives pour intégrer toute la complexité du monde (Simon, 1990) et, d'autre part, les informations sur lesquelles nous raisonnons nous sont généralement transmises par un média social comme nous l'avons vu.

La première tâche de l'individu est donc de déterminer les attentes de son interlocuteur et les raisons pour lesquelles il lui fournit un agencement particulier de conditions vis-à-vis d'un dommage (i.e., modèle causal). Comme le soulignent Hilton et al. (2017), le principe de coopération dans la communication implique de supposer que toute information transmise par notre interlocuteur est pertinente pour la tâche en cours, sinon ces informations n'auraient pas été mentionnées inutilement. Nous pouvons alors nous poser la question de savoir si les scénarios présentés par l'expérimentateur, et plus précisément leur modèle causal, pourraient à ce titre influencer les inférences des participants sur le sens de la tâche qui leur est proposée, notamment en termes d'enquête explicative ou attributive. Pour étudier cela, nous avons

élaboré une première tentative de classification des modèles causaux possibles dans les études sur l'effet de la valence morale. Ce travail d'analyse nous a permis d'identifier quatre modèles causaux distincts, construits en se centrant sur le lien causal entre l'agent et le dommage (voir Figure 13 ci-dessous).

Figure 13

Classification des modèles causaux des scénarios impliquant un agent humain.



Note. La présence d'une flèche permet d'indiquer un lien causal ainsi que son sens. La taille de la flèche permet d'indiquer la force de chaque relation. Un rond vide représente les autres éléments du scénario qui ont contribué au dommage.

Notre hypothèse est que, dans les recherches conduites jusqu'ici concernant l'effet de la valence morale sur le jugement causal, certains modèles causaux (notamment 1 et 2) devraient davantage avoir incité les participants à interpréter la question en termes d'enquête attributive. Les résultats soulignant la prise en compte de la valence morale de l'agent dans le jugement causal, observés dans les recherches ayant utilisé de tels modèles, présenteraient alors un problème d'ordre méthodologique. En effet, une prémisse des chercheurs est que la question de jugement causal mesure la représentation d'un lien causal au sens strict des participants. En effet, ce qui intéresse les chercheurs sur le jugement causal est l'enquête explicative, qui ne concerne que la causalité au sens strict. Or, si la question est interprétée comme une enquête attributive, ce sont les jugements de responsabilité ou de blâme qui sont mesurés. Avant de

détailler les implications pragmatiques de chaque modèle causal, nous avons souhaité savoir quels étaient les modèles les plus utilisés dans la littérature. Si la majorité des études a utilisé des modèles causaux incitant les participants à interpréter la question en termes d'enquête attributive, cela pourrait appuyer l'hypothèse pragmatique de l'effet de la valence morale. Pour ce faire, nous avons réalisé une revue narrative⁴⁴.

Revue narrative de la littérature

La sélection des articles a été faite sur la base de 2 critères d'inclusion. Le premier était la manipulation d'une variable connue pour influencer la valence morale d'un agent humain impliqué dans un événement à l'issue négative (meurtre, vol, etc.) dans au moins une des recherches décrites. Cette VI pouvait renvoyer aux intentions, aux motivations, au comportement, à la personnalité de l'agent, ou encore à la personnalité de la victime. Le second critère était le jugement causal utilisé comme variable dépendante.

Nous avons sélectionné les articles en utilisant une recherche électronique (bases de données PsychINFO, PsychARTICLES et EsbcoHost), complétée d'une analyse manuelle. La recherche dans les bases de données a été effectuée avec les mots clés suivants : « blame-validation », « causation », « cause », « blame », « culpability », « Culpable Control Model », « causal reasoning », « causal judgment », « causal selection », « norm violation ». Nous avons complété le recueil par les articles des principaux auteurs du domaine qui étaient pertinents mais non identifiés électroniquement. Étant donné le caractère inclusif de certains termes (e.g., « cause », « causal judgment »), nous avons obtenu 3228 résultats. Sur ces 3228 résultats obtenus, 24 articles respectant nos critères d'inclusion ont été retenus pour un total de 46 études expérimentales. Ces articles s'étendent sur une période allant de 1981 à 2021. Si cette revue

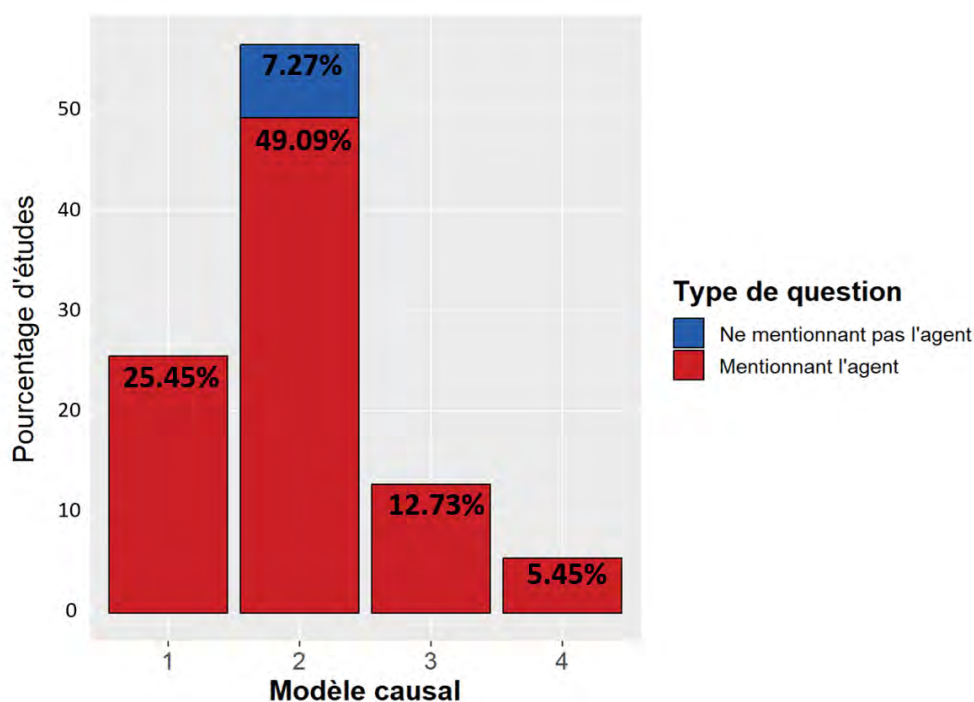
⁴⁴ Le recensement des articles a été réalisé en collaboration avec Jasmine Thrierr (étudiante en Master 2 recherche à l'Université Toulouse-Jean Jaurès) dans le cadre d'un stage de découverte de la recherche au laboratoire CLLE (UMR 5263) entre le 25/02/2022 et le 11/05/2023.

n'est pas exhaustive, nous pouvons toutefois prétendre qu'une part significative des études internationales publiées dans des revues à comité de lecture concernant l'effet de la valence morale sur le jugement causal est présente dans cette revue (pour le détail de chaque étude ; voir Annexe E).

Comme l'illustre la Figure 14 (ci-après), 56 % des modèles causaux utilisés correspondent au modèle 2 (i.e., aucune condition ne se distingue en termes de force causale). Le modèle 1 (lien agent/dommage le plus fort) arrive en seconde position (25.5 %). Les modèles 3 (lien agent/dommage faible) et 4 (lien agent/dommage absent), quant à eux, ont peu été utilisés dans les recherches (13 % et 5.5 % respectivement). Concernant le type de question posée, 92.7% des études ont utilisé une tâche de jugement mobilisant une question mentionnant l'agent. Nous allons désormais proposer une analyse pragmatique de l'utilisation de ces modèles causaux en lien avec la focalisation de la question.

Figure 14

Proportions (%) d'études sur l'effet de la valence morale en fonction du modèle causal utilisé et du type de question posée.



Analyse pragmatique des modèles causaux et de la focalisation de la question

Avant de procéder à une analyse pragmatique de chaque modèle causal, il convient d'évoquer brièvement les raisons méthodologiques, qui selon nous, peuvent rendre compte de la plus forte utilisation tant des questions mentionnant l'agent (vs. ne mentionnant pas l'agent) que du modèle causal 2.

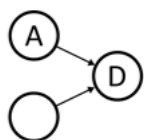
L'utilisation d'une question mentionnant l'agent correspond majoritairement à une question fermée à échelle de réponse permettant d'obtenir un score de jugement causal⁴⁵. L'utilisation d'un score permet ainsi d'augmenter la capacité de la mesure à identifier toute influence de la valence morale de l'agent sur le jugement causal (Alicke, 1992 ; Lagnado & Channon, 2008). En effet, une question ouverte ne permet qu'une mesure dichotomique du jugement causal : l'agent est mentionné ou non.

Néanmoins, concernant l'utilisation d'une question mentionnant l'agent (e.g., Dans quelle mesure l'agent est-il la cause du dommage ?), nous avons déjà vu avec les travaux de Samland et Waldmann (2014, 2015, 2016) qu'il s'agirait d'une mesure incitant les participants à interpréter la question de causalité en termes de responsabilité morale puisqu'elle se focalise sur un facteur ayant une valence morale, à savoir un agent humain. En effet, les participants peuvent inférer que le renvoi à l'agent qui enfreint la norme est intentionnel, et que l'expérimentateur leur demande d'effectuer un jugement de responsabilité. A l'inverse, l'utilisation d'une question ne mentionnant pas l'agent telle qu'une question ouverte (e.g., « Quelle est la cause principale du dommage ? »), ne devrait pas guider vers l'une ou l'autre des deux types d'enquête (explicative, attributive) en raison de la polysémie du mot *cause*. Comme le montre la Figure 13, les seules études ayant utilisé ce type de mesure ont

⁴⁵ Il existe certaines exceptions, notamment l'étude 2 de Samland et Waldmann (2015) que nous avons présentée. Les modalités de réponse mentionnaient notamment l'agent mais il ne s'agissait pas d'un score.

utilisé un récit correspondant au modèle 2 (soit 4 études sur 46). Nous y retrouvons la première étude sur la validation du blâme (Alicke, 1992, Étude 1) dans laquelle la question causale était « Complétez la phrase suivante : La cause principale de cet accident était ___ » (p.369). Les trois autres études ont été menées par Samland et Waldmann (2015, 2016), lesquels n'ont pas répliqué l'effet de la valence morale sur le jugement causal lorsque les modalités de réponse ne mentionnaient pas des agents humains mais des objets physiques.

Bien que cet argument ne soit pas explicitement mis en avant dans la littérature, nous supposons que l'utilisation majoritaire du modèle 2 permet un contrôle expérimental qu'il serait plus difficile d'atteindre avec les autres modèles. Rappelons qu'il s'agit d'un modèle causal dans lequel tous les facteurs présentés contribuent de manière équivalente au dommage. Le scénario le plus célèbre et le plus utilisé est celui de la *pen vignette* que nous avons utilisé comme fil rouge dans le présent manuscrit (Knobe & Fraser, 2008). Pour rappel, un assistant administratif (autorisé à prendre des stylos) et un professeur (non autorisé à prendre des stylos) prennent chacun un stylo et provoquent une pénurie. Le choix de ce modèle peut possiblement s'expliquer par une volonté de rigueur méthodologique de la part des chercheurs : permettre la variation de l'infraction d'une norme entre les deux agents en maintenant constantes toutes les autres conditions du scénario, y compris la force du lien causal entre les conditions. De ce point de vue, le choix des auteurs est tout à fait pertinent. Il peut cependant guider, selon nous, les inférences conversationnelles des participants vers une interprétation de la tâche en termes d'enquête attributive.



Modèle 2

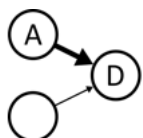
Nous considérons que la configuration « **modèle causal 2** + question ouverte » pose un problème pragmatique. En effet, sachant que chaque condition du récit (e.g., l'assistant administratif et le Professeur) possède une force causale

équivalente vis-à-vis du dommage, il n'y a ici aucune raison logique à sélectionner une

condition comme cause principale plutôt qu'une autre (pour une discussion philosophique sur le sujet, Lewis, 1973 ; Mill, 1843). Si, à la question « Quelle est la cause principale du dommage ? », le mot *cause* renvoie à un lien causal au sens strict, il n'y a donc pas de bonne réponse à la tâche de jugement causal posée aux participants. Le modèle causal rationnel de la situation permet de désigner les liens causaux (i.e., le professeur et l'assistant administratif ont tous les deux causé l'absence de stylos), mais ne permet pas de distinguer une condition en tant que cause principale, car aucune ne présente une force causale supérieure à l'autre. Dès lors, conformément à la présomption de pertinence évoquée précédemment (Sperber & Wilson, 1996), en demandant aux participants de répondre à cette question insoluble, nous ne pouvons écarter l'hypothèse que ceux-ci interprètent le mot *cause* en termes de responsabilité (i.e., *enquête attributive*). En effet, seule cette interprétation permettrait aux participants de supposer que l'expérimentateur leur a posé une question pertinente et de lui fournir, en retour, une réponse pertinente. C'est pourquoi, les participants pourraient désigner le protagoniste méritant d'être sanctionné (i.e., le professeur qui enfreint le règlement) comme responsable du dommage. Cette interprétation pourrait être encore plus saillante lorsque la tâche de jugement causal mobilise une question centrée sur l'agent (49.09 % des études menées et s'inscrivant dans le modèle 2). Cependant, cette désignation n'implique pas nécessairement que leur représentation du modèle causal de la situation soit affectée par une volonté de blâmer le professeur comme le prédirait le CCM. L'effet de la valence morale pourrait davantage être ici la conséquence d'une simple interprétation de la question de causalité posée par l'interlocuteur, relativement à une représentation biaisée du modèle causal.

Un biais étant un écart par rapport à un modèle normatif, encore faut-il que le modèle normatif en question ait une réponse à fournir pour indiquer une cause principale. Il semblerait que ce principe ne puisse être garanti dans les études ayant utilisé le modèle causal 2 et une question ouverte au sein de la tâche de sélection causale. Par conséquent, il ne nous paraît pas

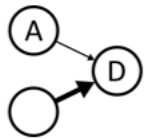
possible d'affirmer que, dans ces études, le jugement causal des participants était biaisé car il n'y a pas de point de référence considéré comme rationnel à la question de causalité. Une solution possible serait de proposer une question ouverte en précisant que plusieurs causes peuvent être mentionnées : « Quelle(s) sont la ou les cause(s) du dommage ? » sans proposer d'options de réponse. Ce type de mesure ne semble pas avoir été testé à ce jour.



Modèle 1

Le deuxième modèle le plus utilisé, à savoir le **modèle causal 1** (utilisé dans 25.5% des recherches passées), pourrait, dans certaines conditions, également inciter les individus à interpréter la tâche comme une *enquête attributive*. Pour rappel, ce modèle correspond à un scénario dans lequel l'agent possède le lien causal le plus fort vis-à-vis du dommage. Dans certains scénarios, ses actions constituent d'ailleurs la seule condition mise en avant. Par exemple, le scénario de l'Étude 3 de Phillips et ses collègues (2015) présente un médecin qui débranche un patient sans-abri, n'ayant plus que deux semaines à vivre, pour abrégé ses souffrances (motivations louables) ou par mépris de classe (motivations blâmables). Ici, le scénario lui-même se focalise sur un facteur unique (l'action du médecin) possédant une valence morale. Aucune autre condition contribuant au décès du patient tel qu'il est survenu n'est mentionnée dans le scénario. A l'instar des questions, les scénarios focalisés sur l'agent devraient, selon nous, orienter l'interprétation de la tâche. Cette focalisation du scénario sur l'agent (malveillant/bienveillant) pourrait conduire les participants à interpréter la tâche de jugement causal qui suit la lecture du scénario comme une tâche de jugement moral ou d'attribution de responsabilité vis-à-vis de l'agent. Contrairement au modèle causal 2, le modèle 1 contient une réponse pertinente à fournir à une question de causalité. Le problème est qu'en absence d'autre condition à mentionner, cette réponse pourrait sembler trop évidente, contrairement à la question de savoir si le médecin peut être considéré comme moralement responsable de la mort de son patient condamné. Conformément à la présomption de pertinence, il paraît plus judicieux ici d'interpréter la tâche comme une enquête

attributive (i.e., évaluer la responsabilité du médecin) plutôt que comme une enquête explicative visiblement trop évidente⁴⁶.

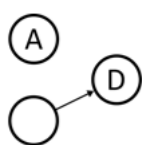


Modèle 3

Ces problèmes d'ordre pragmatiques semblent être moins présents dans le **modèle causal 3**, pourtant l'un des moins utilisés dans la littérature. Dans la mesure où il se compose d'une condition pertinente pouvant être avancée à une réponse de causalité au sens strict et que cette condition n'est pas l'agent, ce modèle 3 oriente probablement moins les participants vers une interprétation de la tâche de jugement causal en termes d'*enquête attributive*. Il y a en effet une réponse pertinente à fournir du point de vue de la causalité au sens strict et celle-ci n'est pas l'agent. Le scénario de Barbara utilisé dans nos deux premières études, lorsque le lien causal entre ses actions et la mort de son mari est faible (i.e., le poison n'est pas assez fort mais laisse un goût désagréable qui incite le mari à changer de plat), est un exemple de scénario entrant dans le modèle causal 3. Dans le cadre d'une *enquête explicative*, interrogés sur la causalité au sens strict, les participants devraient avancer la condition la moins sensible aux circonstances, à savoir « la réaction allergique ». Dans le cadre d'une *enquête attributive*, où la causalité peut être interprétée en termes de responsabilité morale, la réponse des participants pourrait être « Barbara » si ses motivations sont blâmables (i.e., hériter de l'argent de son mari) et « Jean » si celles-ci sont louables (i.e., mari violent). Les deux interprétations (*enquêtes explicative/attributive*) sont toujours possibles, chacune ayant sa réponse pertinente, mais ce modèle ne semble pas inciter particulièrement les participants à privilégier une interprétation de la causalité au détriment d'une autre. Pour les aider à interpréter la demande faite par l'interlocuteur, ils pourront cependant se servir d'autres indices, tels que la formulation de la question.

⁴⁶ Le texte évoque explicitement que c'est l'action du médecin qui l'a débranché et qui a provoqué sa mort : « Then, participants read that Dr. Bennett proceeded to disconnect the man from the machine, and the man quickly dies. » (Phillips et al, 2015, p.36)

Dans le modèle causal 3, les questions centrées sur l'agent pourraient orienter l'interprétation de la causalité vers la recherche d'un responsable comme nous l'avons vu, et ainsi favoriser l'effet de la valence morale sur le jugement causal. En revanche, l'utilisation d'une question ouverte ne mentionnant pas l'agent ne devrait pas favoriser l'une ou l'autre des interprétations. Néanmoins, les résultats de l'Étude 1 de cette thèse semblent suggérer que la question était principalement interprétée en termes de causalité au sens strict (l'allergie était majoritairement désignée comme cause dans la condition *lien causal faible*, l'agent n'était que marginalement sélectionné).



Concernant le **modèle 4**, dans lequel l'agent n'est pas causalement lié au dommage (e.g., lien causal préempté), les ambiguïtés pragmatiques ne se posent pas puisque l'agent ne devrait pas être désigné comme cause, peu importe le type d'enquête. En effet, un lien causal est requis tant pour le jugement de causalité au sens strict (cf. Chapitre 3 de cette thèse, Études 1 à 4) que pour l'attribution d'une responsabilité (cf. Chapitre 3, Études 3 et 4).

En résumé, nous suggérons qu'une grande partie des récits et des questions utilisées dans la littérature concernant l'effet de la valence morale sur le jugement causal inciterait les participants à inférer que la tâche de jugement causal, qui leur est proposée, s'inscrit dans le cadre d'une *enquête attributive*, c'est-à-dire la recherche d'un responsable. Cette observation est particulièrement vraie pour les recherches ayant recours à un scénario entrant dans le modèle causal 2, et pour lesquelles la tâche de jugement causal se traduit par une question mentionnant l'agent. La première raison à cela est que ce modèle causal ne permet pas de sélectionner une cause principale dans le cadre d'une *enquête explicative*. La deuxième raison est qu'une question centrée sur l'agent pourrait inciter les participants à interpréter la tâche comme un jugement de responsabilité (cf. Samland & Waldmann, 2014, 2015, 2016). À l'inverse, notre hypothèse est que la configuration la moins favorable à une interprétation en

termes de responsabilité est l'utilisation du modèle 3 (un autre facteur que l'agent se distingue en termes de force causale) et d'une question ouverte (qui ne mentionne pas spécifiquement l'agent). Cette opérationnalisation ne semble jamais avoir été testée dans la littérature. Seule l'Étude 1 du présent manuscrit (condition *lien faible*) aurait testé cette configuration. Nous utiliserons donc ce type de modèle et de question pour nos prochaines études afin de neutraliser au mieux les variables pouvant affecter l'interprétation du mot *cause* par les participants.

Effet du cadrage conversationnel de la tâche sur le jugement causal

Les travaux de Samland et Waldmann (2014, 2015, 2016) sur le cadrage conversationnel et l'effet de la valence morale ont uniquement utilisé des scénarios présentant un modèle causal 2, qu'ils nomment structure causale conjonctive (i.e., deux agents contribuent de manière équivalente au dommage), et une tâche de jugement causal mobilisant une question mentionnant l'agent. Nous avons vu les limites pragmatiques susceptibles d'accompagner l'utilisation d'un tel modèle causal et d'une question de causalité de cette nature. Pour répondre à ces limites, il conviendrait de tester le cadrage conversationnel avec un modèle causal 3. L'agent y est le facteur le plus pertinent à mentionner si la question de causalité peut être interprétée comme renvoyant à la responsabilité mais pas si elle renvoie à la causalité au sens strict). Par ailleurs, comme nous l'avons proposé, l'utilisation d'une question ouverte induirait le moins d'inférences conversationnelles chez les participants. Par conséquent, dans les études 5 et 6 présentées ci-après, nous avons testé l'influence du cadrage conversationnel sur l'effet de la valence morale à travers l'utilisation d'un récit correspondant au modèle causal 3, et d'une tâche de jugement causal mobilisant une question ouverte qui, de fait, ne mentionne pas d'agent humain.

Les hypothèses seront très différentes concernant l'approche théorique privilégiée. D'un côté, les prédictions issues du CCM seraient la réplique de l'effet de la valence morale

indépendamment du cadrage de la question, puisque c'est la représentation du lien causal qui est biaisée. Or, si le lien causal entre l'agent blâmable et le dommage est surestimé, alors les participants désigneraient davantage l'agent blâmable (vs. louable) comme la cause, cela que la tâche de jugement causal utilise une question renvoyant à la causalité au sens strict ou une question renvoyant à la responsabilité. En revanche, conformément à l'approche pragmatique, nous pouvons supposer que les motivations de l'agent n'affecteraient pas la représentation du lien causal, mais influencerait plutôt la sélection causale, c'est-à-dire le choix de la condition la plus pertinente à mentionner à l'interlocuteur. Conformément à la *maxime de relation* (Grice, 1975 ; ou *principe de pertinence*, Sperber & Wilson, 1996), si la question est interprétée comme une recherche de responsable (i.e., *enquête attributive*), les participants désigneraient davantage l'agent aux motivations blâmables (vs. louables). En revanche, si la question est interprétée comme une demande d'explication de l'événement (i.e., *enquête explicative*), alors ils devraient désigner le facteur le moins sensible aux circonstances de fond (i.e., ayant la plus grande force causale) et l'effet des motivations de l'agent sur le jugement causal ne devrait pas être observé. En somme, selon cette approche, les états mentaux de l'agent ne devraient être pris en compte que lorsque le contexte de la tâche de jugement (i.e., cadrage conversationnel) oriente l'interprétation de la causalité vers une *enquête attributive*. Au vu des résultats de Samland et Waldmann (2014, 2015, 2016), de nos résultats différents en fonction du type de mesure utilisée (Étude 1 vs. Étude 2 à 4) et de notre analyse de la revue narrative de la littérature, proposant que les études incitaient majoritairement les participants à interpréter la tâche comme une enquête attributive, nous soutenons l'hypothèse pragmatique.

Étude 5 : Influence du cadrage conversationnel sur l'effet de la valence morale de l'agent sur le jugement causal

Méthode

Design. Le plan expérimental était un plan factoriel 2 (Motivations : louables, blâmables) × 2 (Cadrage conversationnel : enquête explicative, enquête attributive) en intersujets.

Participants. Nous avons recruté 147 participants (82.31% de femmes, $n = 121$) par le biais du réseau social Facebook (soit environ 35 participants par condition). La moyenne d'âge était de 27.30 ans ($SD = 8.75$; min = 18 ; max = 59). A nouveau, aucune compensation financière n'a été versée aux participants et le protocole a été approuvé par le Comité d'Éthique de la Recherche de Toulouse.

Matériel et Procédure. En premier lieu, les participants étaient invités à lire le formulaire de consentement et à indiquer leur accord s'ils l'approuvaient. Une fois leur accord fourni, les participants ont renseigné des informations sociodémographiques (genre, âge, niveau en français⁴⁷), puis il leur était demandé de lire attentivement l'un des 4 scénarios distribués aléatoirement. Nous avons adapté le scénario de l'avalanche de Hilton et ses collègues (2010) en modèle causal 3 avec un dommage létal en manipulant les motivations de l'agent (skieur) :

« Cet hiver, la neige s'est accumulée créant de nombreuses plaques de neige dans les montagnes. Un employé de mairie était en train de finir de poser les balises de hors-piste lorsqu'il se fractura la jambe et perdit connaissance. Un skieur aperçut l'employé de mairie blessé au loin. Malgré le panneau indiquant qu'il était interdit de quitter le circuit régulier, il

⁴⁷ La présente étude portant sur la pragmatique du langage, nous avons décidé d'inclure uniquement les participants ayant un niveau C2 minimum en français.

partit en hors-piste dans le but de porter secours à (motivations louables) /de dérober les affaires de (motivations blâmables) l'employé blessé et inconscient. Suite à sa présence sur cette partie de la montagne, une grande plaque de neige se détacha la faisant rouler sur la pente provoquant une importante avalanche qui détruisit une partie du village. Cet accident engendra le décès de plusieurs habitants. »

Les caractéristiques de ce scénario répondaient à celles du modèle causal 3 car, l'agent n'était pas la condition ayant le lien causal le plus fort avec le dommage puisque ce lien dépendait du détachement de la plaque de neige (le hors-piste du skieur n'aurait pas provoqué la mort des habitants si la plaque ne s'était pas détachée). La plaque de neige correspondait au facteur ayant le lien causal le plus fort (i.e., le moins sensible aux circonstances) avec le dommage dans ce scénario. De plus, comme dans les Études 3 et 4, l'état épistémique de l'agent vis-à-vis du dommage était maintenu constant (i.e., l'agent n'avait pas l'intention de détruire le village quelles qu'étaient ses motivations à faire du hors-piste).

Après avoir lu le scénario, les participants étaient exposés à l'une des deux conditions de cadrage conversationnel (*enquête explicative*, *enquête attributive*). Nous avons utilisé la description d'un interlocuteur fictif en manipulant son statut (journaliste, juge) et son objectif (comprendre l'événement vs. trouver un responsable à sanctionner). Dans la condition *enquête explicative*, les participants étaient invités à lire l'énoncé suivant :

« Un journaliste reporter arrive sur place et souhaite comprendre ce qui a causé le décès des habitants du village. Pour le moment, il n'a aucune information sur ce qu'il s'est passé. Vous avez assisté à toute la scène, il vous demande de lui indiquer la cause principale du décès des habitants du village tel qu'il est survenu. ».

Dans la condition *enquête attributive*, l'énoncé était celui-ci :

« Un juge engage une procédure pour déterminer s'il existe un responsable du décès des habitants du village pour attribuer d'éventuelles sanctions. Pour le moment, il n'a aucune information sur ce qu'il s'est passé. Vous avez assisté à toute la scène, il vous demande de lui indiquer la cause principale du décès des habitants du village tel qu'il est survenu. ».

Après avoir lu le récit et le cadrage conversationnel, la tâche de jugement causal était présentée aux participants des deux conditions sous la forme suivante : « La cause principale du décès des habitants est _____ ».

Dans la mesure où, comme le soulignent Lagnado et Channon (2008), les questions ouvertes sont des mesures moins fines que les questions à échelle centrées sur l'agent, nous avons également ajouté des questions qui permettraient de mesurer les jugements de causalité au sens strict sans mentionner le mot *cause* et provoquer d'ambiguïté conversationnelle (pour une analyse similaire, voir Samland & Waldmann, 2016, Étude 1). Il s'agissait de la mesure Delta-P (ΔP) initialement élaborée par Cheng et Novick (1990, 1992). Pour calculer ce score, les participants devaient estimer le risque que le dommage survienne étant donné la présence d'une condition (e.g., « Veuillez estimer, à partir du moment où le skieur fait du hors-piste, le risque que les habitants décèdent ? ») et le risque que le dommage survienne étant donné l'absence de cette même condition (e.g., « Veuillez estimer, dans le cas où le skieur ne fait pas de hors-piste, le risque que les habitants décèdent ? »), et ce, pour chaque condition de l'événement (i.e., accumulation de la neige dans les montagnes, hors-piste du skieur, blessure de l'employé, plaque de neige, scène se déroulant en hiver). Les participants répondaient à ces questions sur une échelle de 0 (Impossible) à 10 (Certain). Ces questions permettaient de calculer le score ΔP de chaque condition du scénario. Celui-ci était calculé à partir de la différence entre la situation hypothétique dans laquelle la condition est présente et celle dans

laquelle elle est absente⁴⁸. Conformément au critère de sensibilité, nous nous attendions à ce que le facteur le moins sensible aux circonstances (i.e., la plaque de neige) reçoive le score ΔP le plus élevé, qualifiant ainsi la condition ayant la plus grande force causale. Conformément à l'approche pragmatique et aux résultats de Samland et Waldmann (2016, Étude 1), nous nous attendions à ce que le score ΔP de l'agent ne soit pas affecté par ses motivations (louables/blâmables).

Enfin, nous avons demandé aux participants d'indiquer leur sentiment vis-à-vis du skieur sur une échelle de 0 à 10 (où 0 = Très négatif ; 5 = neutre ; 10 = Très positif). Un test *t* à mesures indépendantes a mis en évidence que les participants affectés à la condition *blâmables* ont évalué l'agent plus négativement ($M = 2.21$, $Mdn = 1$, $SD = 2.38$) que les participants affectés à la condition *louables* ($M = 6.90$, $Mdn = 7$, $SD = 1.74$), $t(118) = 12.4$, $p < .001$, $d = 2.27$, $ICd\ 95\% = [1.71, 2.82]$. Les distributions des scores d'évaluations sont disponibles en Annexe F.

Codage et notation des données. Comme pour l'Étude 1 (pp.79-80), nous avons créé deux variables à partir des réponses sur la cause principale désignée. Dans un premier temps, nous avons classé les réponses des participants en 5 catégories : (1) Plaque ; (2) Skieur ; (3) Neige ; (4) Blessure de l'employé ; (5) Autre. Ensuite, afin d'évaluer si l'agent était davantage sélectionné que tout autre facteur, les réponses étaient classées en deux catégories : celles mentionnant le skieur (e.g., « le skieur », « la plaque et le skieur ») et celles ne le mentionnant pas (e.g., « l'accumulation de neige », « la blessure de l'employé de mairie »).

Résultats

Sélection causale. Dans un premier temps, nous avons souhaité savoir quelle cause était majoritairement avancée par les participants pour chaque cadrage conversationnel

⁴⁸ $\Delta P = P(D|C) - P(D|C^-)$ où D = Dommage et C = Condition

(enquête explicative, enquête attributive). Pour ce faire, nous avons conduit des tests du Khi-deux d'ajustement sur les réponses fournies par les participants pour chaque cadrage conversationnel (α corrigé à .025). Comme indiqué dans le Tableau 11 (ci-dessous), la distribution des causes sélectionnées n'était pas uniforme, cela quel que soit le cadrage conversationnel. La condition la plus désignée comme cause principale du décès des habitants par les participants était la plaque de neige qui s'est détachée. La plaque de neige était davantage sélectionnée par rapport à la seconde condition la plus sélectionnée, à savoir le skieur, aussi bien dans le cadre d'une enquête explicative ($p < .001$) que d'une enquête attributive ($p < .001$).

Tableau 11

Proportions de sélection des différentes conditions du scénario comme cause en fonction du cadrage conversationnel (enquête explicative, enquête attributive)

		Type de cause sélectionnée					χ^2 (df)	p	Cohen's w
		Plaque de neige	Skieur	Blessure de l'employé	Neige	Autre			
Enquête	Explicative (n = 73)	.712	.219	.014	.027	.027	130.36 (4)	<.001	1.33
	Attributive (n = 74)	.716	.162	.027	.054	.041	127.49 (4)	<.001	1.31
	Total (N = 147)	.714	.191	.020	.041	.003	257 (4)	<.001	1.32

Dans un second temps, l'occurrence de la réponse « le skieur » comme cause de la mort des habitants a été entrée comme variable dépendante dans une régression logistique binomiale, avec le cadrage conversationnel et les motivations comme prédicteurs, ainsi qu'une interaction Motivations \times Enquête. Ce modèle ne prédisait pas les données de manière significative, $\chi^2(3) = 4.77, p = .189$. En effet, les analyses n'ont pas montré d'effet principal significatif des motivations, (OR = 1.77, 95% CI [0.58, 5.59], $p = .318$), ni du cadrage conversationnel, (OR = 1.25, 95% CI [0.43, 3.68], $p = .682$) sur la sélection de l'agent comme cause principale

du dommage. Enfin, nous n'avons pas non plus observé d'effet d'interaction entre ces deux variables, OR = 1.81, 95% CI [0.32, 12.07], $p = .516$.

Comme indiqué dans le Tableau 12 (ci-après), les résultats des analyses post-hoc n'ont pas montré d'effet des motivations sur la sélection causale de l'agent, et ce, quel que soit le cadrage conversationnel (valeurs p supérieures à .05 et intervalles de confiance de l'OR incluant le 1). Les facteurs bayésiens proches de 1 (particulièrement pour l'enquête attributive) semblaient notamment indiquer que les données ne permettaient pas de départager l'hypothèse nulle de l'hypothèse alternative de présence d'un effet des motivations sur la sélection causale (Dienes, 2014 ; Van den Bergh et al., 2020).

Tableau 12

Proportions de sélections de l'agent comme cause principale en fonction de ses motivations et du cadrage conversationnel

		Motivations		OR	p	BF ₁₀
		Louables	Blâmables			
Enquête	Explicative ($n = 73$)	.175	.273	1.77 [0.58, 5.59]	.318	0.39
	Attributive ($n = 74$)	.086	.231	3.20 [0.86, 15.46]	.103	0.82

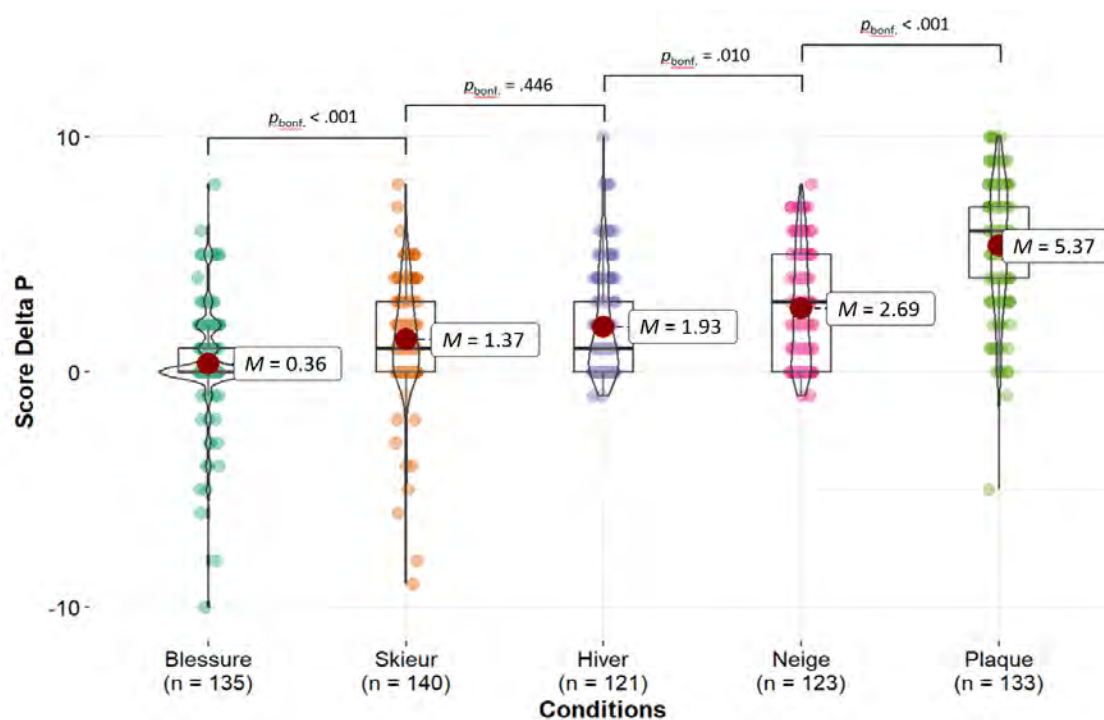
Analyse de sensibilité. Nous avons conduit des analyses de sensibilité avec le logiciel G-Power (version 3.1.9.6) pour estimer quelle taille d'effet minimum nos données permettaient de détecter avec une puissance statistique de 80% (Lakens, 2022). Les analyses ont montré que le plus petit odds ratio que nous pouvions détecter était de 3.64 dans la condition *enquête explicative*, et de 3.61 dans la condition *enquête attributive* ($\beta = .20$; $\alpha = .05$). Nos résultats nous permettaient donc de conclure que si les motivations de l'agent ont réellement un effet sur le jugement causal pour ce scénario, l'odds ratio de cet effet est inférieur à 3.61 (enquête attributive) et 3.64 (enquête explicative). Ces odds ratio correspondent à des tailles d'effets

moyennes (voir Chen et al., 2010). Ainsi, notre étude ne permettait pas de conclure sur la présence ou l'absence d'un effet des motivations sur le jugement causal dans ce scénario si cet effet est de petite taille.

Scores ΔP des conditions. La Figure 15 (ci-après) reprend les scores ΔP de chaque condition ainsi que les tests post-hoc avec correction de Bonferroni. Une ANOVA à mesures répétées a révélé un effet principal du type de condition sur le score ΔP avec une très large taille d'effet, $F(4, 120) = 99.00, p < .001, \eta_p^2 = .452$. Conformément à nos attentes, la plaque de glace était la condition ayant le score ΔP le plus élevé. Ce score était supérieur à celui du skieur, $t(120) = 13.02, p_{\text{bonf.}} < .001$. En revanche, contrairement à l'hypothèse pragmatique, le score ΔP du skieur ayant des motivations blâmables ($M = 1.91, Mdn = 2, SD = 2.04$) était significativement plus élevé que celui du skieur ayant des motivations louables ($M = 0.89, Mdn = 1, SD = 2.62$), $t(138) = 2.53, p = .013, d = 0.43, ICd\ 95\% = [0.09, 0.77], BF_{10} = 3.23$.

Figure 15

Distribution des scores Delta P pour chaque condition du scénario.



Analyses de médiation. Afin de déterminer si l'effet des motivations sur le score ΔP de l'agent pouvait s'expliquer par l'évaluation négative à son encontre (variable médiatrice), une méthode non paramétrique par bootstrap (5000 itérations) a été conduite. Les résultats ne permettaient pas d'indiquer que la relation entre les motivations et le score ΔP s'expliquait par l'évaluation envers l'agent (effet indirect non significatif, voir Tableau 13, ci-dessous). A nouveau, l'absence d'effet direct pouvait suggérer un problème de multicollinéarité (Zhao et al., 2010) lié à l'association très forte entre notre variable indépendante et notre variable médiatrice (cf. la très forte taille d'effet des motivations sur l'évaluation).

Tableau 13

Analyses de médiation par bootstrap de l'effet des motivations de l'agent (variable indépendante) sur le score ΔP du skieur (variable dépendante) médiatisé par l'évaluation de l'agent (variable médiatrice)

Effet	Estimation	Intervalle de confiance (95%) [inférieur, supérieur]	<i>p</i>
Direct	0.398	[-0.96, 1.81]	.580
Indirect	0.789	[-0.15, 1.76]	.101
Total	1.187	[0.31, 2.05]	.009

Discussion de l'étude 5

L'objectif de cette étude était de confronter les hypothèses du CCM et de l'approche pragmatique en manipulant le cadrage conversationnel entourant la question de jugement causal. Les deux approches prédisaient un effet de la valence morale de l'agent sur le jugement causal dans la condition de l'*enquête attributive*. Concernant l'*enquête explicative*, si le CCM prédisait également la réplification de l'effet, l'approche pragmatique prédisait au contraire une neutralisation de l'effet.

Étudions pour commencer notre hypothèse pragmatique concernant le type d'enquête. Pour rappel, cette hypothèse est que l'effet de la valence morale n'est pas lié à une surestimation du lien causal entre l'agent et le dommage, mais à une interprétation de la question de causalité en termes d'*enquête attributive* (i.e., recherche d'un responsable). Dans la condition *enquête explicative*, les résultats obtenus corroborent notre hypothèse, à savoir que la condition la moins sensible aux circonstances de fond (i.e., la plaque de neige) était la condition la plus désignée par les participants. Il s'agissait également de la condition ayant reçu le score ΔP (i.e., utilisée pour mesurer le jugement de causalité au sens strict) le plus élevé. Dans la condition *enquête attributive* censée être interprétée en termes de recherche d'un responsable (i.e., un juge qui cherche un responsable à sanctionner), les résultats observés étaient inconsistants avec notre hypothèse car, là aussi, la plaque de glace était davantage mentionnée que l'agent. Si ces résultats questionnent l'hypothèse pragmatique du jugement causal, ils offrent toutefois de nouveaux arguments en faveur de la connaissance et de la mobilisation prioritaire des critères normatifs du jugement causal par les individus, notamment de la sensibilité des conditions. Ceci vient corroborer l'hypothèse selon laquelle les individus sont capables d'évaluer la condition la moins sensible aux circonstances de fond et de la désigner comme cause principale pour expliquer un dommage à un interlocuteur. A l'instar de toutes nos études jusqu'à présent, ces résultats corroborent l'hypothèse de la connaissance et de la mobilisation.

Voyons à présent l'effet de la valence morale de l'agent sur sa sélection causale. Nos résultats n'ont pas permis de mettre en évidence un effet de la valence morale de l'agent sur sa sélection en tant que cause principale et ce, quel que soit le cadrage conversationnel (*enquête explicative vs. attributive*). Bien que l'effet de la valence morale n'ait pas été répliqué pour la sélection causale, nous avons retrouvé un effet des motivations de l'agent sur son score ΔP . Ces résultats viennent à l'encontre de ceux de Samland et Waldmann (2016) et corroborent

davantage les prédictions du CCM. Selon lesquelles, pour rappel, l'estimation du lien causal au sens strict entre l'agent et le dommage est biaisée par une évaluation négative à l'égard de l'agent. Nous voyons deux explications possibles à cette divergence de résultats entre les études de Samland et Waldmann et la nôtre. La première résiderait dans la sévérité du dommage comme l'a montré l'Étude 4 du présent travail de thèse. Nous avons utilisé un scénario avec un dommage légal, alors que leurs études présentaient des dommages de faible gravité (i.e., une pénurie de stylos et une panne d'ordinateur). Cette interprétation irait dans le sens du CCM : l'évaluation du lien causal entre l'agent et le dommage n'est biaisée que lorsque l'évaluation spontanée négative est suffisamment importante envers l'agent (Alicke et al., 2011, p.9). Or, la sévérité du dommage de notre étude engendrerait des évaluations spontanées négatives plus importantes que dans les études des deux auteurs. La seconde explication est la suivante : si nous reprenons l'analyse pragmatique proposée dans notre revue de littérature, la mesure ΔP est une question mentionnant un agent humain (i.e., Veuillez estimer, à partir du moment où le skieur fait du hors-piste, le risque que les habitants décèdent ?). Il est donc possible que cette question ait été interprétée comme une enquête attributive. Cette interprétation serait soutenue par l'absence de réplification de l'effet de la valence morale avec une question ouverte dans cette même étude. Si cette analyse est correcte, l'ambiguïté des tâches de jugement causal ne reposerait pas seulement sur la polysémie du mot *cause*, contrairement à ce qu'avancent Samland et Waldmann. Les mesures du jugement causal en général pourraient être sujettes à deux interprétations possibles (*explicative* vs. *attributive*). De futures études pourraient étendre l'étude de l'ambiguïté pragmatique des tâches de jugement causal au-delà du seul mot *cause* et investiguer une potentielle ambiguïté conversationnelle de la mesure ΔP .

La non-réplification de l'effet de la valence morale sur la sélection causale reste toutefois surprenante. Elle ne semble pas s'expliquer par un problème d'opérationnalisation des motivations de l'agent puisque nous avons obtenu une très large taille d'effet des motivations

sur l'évaluation de l'agent par les participants ($d = 2.27$). Comme nous l'avons évoqué, le CCM comme la pragmatique prédisaient une répliation de l'effet de la valence morale dans la condition de l'*enquête attributive*. Cette contradiction nous a amené à nous interroger sur l'opérationnalisation de cette condition. Nous avons fait le choix de faire varier le statut de l'interlocuteur fictif (juge ou journaliste), ainsi que ses objectifs (comprendre ou trouver un responsable à punir). Il s'agissait de mobiliser la maxime de relation (Grice, 1975) stipulant que le répondant veille à être pertinent vis-à-vis des objectifs de l'interlocuteur (i.e., désigner un responsable au juge et désigner la condition la plus informative pour comprendre l'événement au journaliste). Or, dans la condition *enquête attributive*, nous avons précisé que le juge n'avait aucune information sur l'événement. L'absence de connaissances de l'interlocuteur sur un événement pourrait donc avoir conduit le répondant à inférer que la question causale visait dans un premier temps à lui expliquer ce qu'il s'est passé (i.e., causalité au sens strict). Ceci renvoie à la maxime de quantité (Grice, 1975) stipulant qu'il faut fournir suffisamment d'informations à son interlocuteur (i.e., désigner l'événement le plus informatif au juge dans un premier temps pour lui expliquer l'événement). L'*enquête attributive* (i.e., désigner un responsable) ne débiterait que lorsque l'*enquête explicative* est arrivée à son terme et a permis à l'ensemble des interlocuteurs de comprendre l'événement. C'est ce qu'ont montré Hilton et ses collègues (1992) en analysant les articles du New-York Times après l'accident de la navette spatiale Challenger. Dans un premier temps, le média cherchait à comprendre l'événement en se focalisant sur la rupture des joints toriques (i.e., *enquête explicative*). Dans un second temps, il cherchait à désigner les agents humains pouvant être responsables de l'accident, tels que les dirigeants de l'entreprise Thiokol qui a fabriqué les joints (i.e., *enquête attributive*). Par conséquent, pour opérationnaliser plus fidèlement une *enquête attributive*, il semblerait nécessaire que l'interlocuteur fictif sache déjà ce qu'il s'est passé et que les participants n'aient donc pas à lui expliquer l'événement en premier lieu.

Étude 6 : Influence du cadrage conversationnel sur l'effet de la valence morale de l'agent sur le jugement causal (nouvelle opérationnalisation du cadrage)

Nous avons donc souhaité répliquer l'étude 5 en précisant que l'interlocuteur fictif connaissait déjà l'événement dans la condition *enquête attributive*. Par ailleurs, nous avons augmenté la taille de l'échantillon, afin de pouvoir capter une taille d'effet plus faible pour la tâche de sélection causale. Les hypothèses étaient les mêmes que dans l'étude précédente. L'hypothèse issue du CCM prédisait la réplication de l'effet des motivations de l'agent sur sa sélection causale en tant que cause principale dans les deux conditions de cadrage conversationnel (*enquête explicative, enquête attributive*). L'hypothèse issue de l'approche pragmatique, quant à elle, prédisait que cet effet ne devrait être répliqué que dans la condition de l'*enquête attributive* (i.e., un juge cherche un responsable).

Méthode

Design. Le plan expérimental était un plan factoriel 2 (Motivations : louables, blâmables) × 2 (Cadrage conversationnel : enquête explicative, enquête attributive) en inter-sujets.

Participants. Notre échantillon était composé de 260 participants (226 femmes, 28 hommes, 6 autres) recrutés sur le réseau social Facebook (soit environ 65 participants par condition). La moyenne d'âge était de 29.90 ans ($SD = 11.5$). Tous les participants étaient volontaires pour participer à la recherche et ont donné leur consentement éclairé. Avec cette taille d'échantillon, l'analyse de sensibilité a montré que nous pouvions capter une taille d'effet minimale dans chaque condition de $OR = 2.67$ ($\beta = .20$; $\alpha = .05$).

Matériel et Procédure. Le matériel et la procédure très proches de l'étude précédente. Nous avons en effet modifié l'opérationnalisation de la VI cadrage conversationnel. Dans la condition *enquête attributive*, les participants étaient invités à lire l'énoncé suivant :

« Un juge souhaite savoir s'il existe un responsable du décès des habitants. Il a eu accès au rapport des enquêteurs et sait donc ce qu'il s'est passé. Il vous demande de lui indiquer la cause principale du décès des habitants tel qu'il est survenu. ». Ici, il était précisé que l'interlocuteur fictif avait connaissance de la situation.

Dans la condition *enquête explicative*, l'énoncé était celui-ci :

« Un journaliste souhaite comprendre ce qu'il s'est passé. Il n'a aucune information sur l'événement, il sait uniquement que plusieurs habitants sont décédés. Il vous demande de lui indiquer la cause principale du décès des habitants tel qu'il est survenu. ».

De façon semblable à l'étude précédente, les participants complétaient ensuite la phrase suivante : « La cause principale du décès des habitants est _____ ». La question de l'évaluation de l'agent était également identique à l'Étude 5 : « Quelle est votre sentiment vis-à-vis du skieur ? » sur une échelle de 0 à 10 (où 0 = Très négatif ; 5 = neutre ; 10 = Très positif).

Nous avons ajouté deux questions ouvertes qui ne figuraient pas dans l'étude précédente : « Selon vous, quel est l'élément le plus pertinent à sélectionner pour désigner un responsable moral du décès des habitants ? » et « Selon vous, quel est l'élément qui a le plus contribué au décès des habitants ? ». Ces deux questions visaient à confirmer nos attentes vis-à-vis du scénario, à savoir que la plaque de glace devrait être considérée comme l'élément qui a le plus contribué au dommage (i.e., causalité au sens strict) et que le skieur, particulièrement lorsque ses motivations sont blâmables, serait désigné comme l'élément le plus pertinent à désigner comme responsable moral. Enfin, nous avons gardé uniquement les mesures ΔP de 0 (impossible) à 10 (certain)⁴⁹ pour le skieur et la plaque de neige.

⁴⁹ Pour rappel, le score ΔP d'une condition est calculé à partir de la différence entre la situation hypothétique dans laquelle la condition est présente et celle dans laquelle elle est absente : $\Delta P = P(D|C) - P(D|C^-)$ où D = Dommage et C = Condition

Résultats

Analyses préliminaires. Un test *t* à mesures indépendantes a montré que l'agent aux motivations blâmables était évalué plus négativement ($M = 2.10$, $Mdn = 2$, $SD = 1.97$) que l'agent aux motivations louables ($M = 6.45$, $Mdn = 7$, $SD = 1.80$), $t(209) = 16.7$, $p < .001$, $d = 2.31$, $ICd\ 95\% = [1.89, 2.72]$. De plus, comme indiqué dans le Tableau 14 (p.165), le skieur était la condition la plus sélectionnée pour désigner un responsable moral du dommage ($p < .001$). Les participants du scénario *motivations blâmables* désignaient significativement plus le skieur comme responsable moral (32%), que les participants du scénario *motivations louables* (18.5%), $OR = 3.46$, $95\% CI [1.93, 6.19]$, $p < .001$. Concernant le facteur considéré comme ayant le plus contribué au dommage, la plaque de neige était majoritairement désignée ($p < .001$). Contrairement à nos attentes, le skieur était davantage désigné comme le contributeur le plus important vis-à-vis du dommage lorsque ses motivations étaient blâmables (11.9%), plutôt que louables (3.5%), $OR = 2.23$, $95\% CI [1.08, 4.57]$, $p = .029$. Afin d'explorer si l'interprétation de la question du facteur le plus contributeur a été influencée par le cadrage conversationnel, nous avons testé l'interaction entre les motivations et le cadrage conversationnel sur la sélection du skieur comme contributeur. Les analyses n'ont pas montré d'effet d'interaction statistiquement significatif entre le cadrage conversationnel et les motivations de l'agent sur la sélection du contributeur le plus important, $OR = 0.50$, $95\% CI [0.12, 2.14]$, $p = .498$.

Tableau 14

Proportions de sélections de l'agent comme responsable moral et contributeur le plus important en fonction de ses motivations et du cadrage conversationnel

	Plaque de neige	Type de condition sélectionnée					χ^2 (df)	p	Cohen's w
		Skieur	Mairie	Météo	Personne	Autre			
Responsable moral (n = 212)	.060	.505	.130	.055	.040	.210	189.1 (5)	<.001	0.972
Contributeur le plus important (n = 212)	.467	.184	.052	.151	.000	.146	104.7 (4)	<.001	0.703

Sélection causale. Des tests du Khi-deux d'ajustement pour chaque condition de la VI cadrage conversationnel ont montré que la distribution des causes sélectionnées n'était pas homogène (voir Tableau 15, ci-dessous). La plaque de neige était la condition la plus sélectionnée comme cause principale du dommage, devant la sélection du skieur, aussi bien pour l'enquête explicative ($p < .001$) que pour l'enquête attributive ($p < .001$).

Tableau 15

Proportions de sélection des différentes conditions du scénario comme cause principale en fonction du cadrage conversationnel (enquête explicative vs. attributive)

		Type de cause sélectionnée					χ^2 (df)	p	Cohen's w
		Plaque de neige	Skieur	Blessure de l'employé	Neige	Autre			
Enquête	Explicative (n = 130)	.785	.131	.008	.046	.023	365.7 (5)	<.001	1.68
	Attributive (n = 130)	.654	.185	.031	.077	.038	237.6 (5)	<.001	1.35
	Total (N = 260)	.720	.157	.019	.065	.031	599 (5)	<.001	1.32

Nous avons conduit une régression logistique binomiale pour tester les effets du cadrage conversationnel et des motivations de l'agent, ainsi qu'un effet d'interaction entre ces deux variables, sur la probabilité que les participants sélectionnent l'agent ou non. Le modèle

de régression logistique était statistiquement significatif, $\chi^2(3) = 19.8$, $p < .001$. Celui-ci expliquait 9.75% de la variance de sélection de l'agent. Comparativement à l'agent aux motivations louables, celui aux motivations blâmables était quatre fois plus susceptible d'être désigné comme cause principale de l'incident, OR = 4.13, 95% CI [1.35, 15.47], $p = .020$. En revanche, les participants pour qui la question causale était inscrite dans une enquête attributive n'ont pas significativement plus désigné l'agent, que ceux pour qui la question causale était inscrite dans une enquête explicative, OR = 1.33, 95% CI [0.34, 5.59], $p = .681$. Les analyses n'ont pas non plus révélé d'effet d'interaction entre les motivations et le cadrage conversationnel, OR = 1.30, 95% CI [0.25, 6.45], $p = .748$. Comme indiqué dans le Tableau 16 (ci-après), les participants de la condition *motivations blâmables* avançaient plus l'agent comme cause de l'incident, cela pour les deux types d'enquêtes. En revanche, le facteur bayésien pour la condition *enquête explicative* témoignait d'un soutien faible ($BF_{10} < 3$) pour l'hypothèse alternative de présence d'un effet des motivations sur la sélection de l'agent comme cause principale. Ce soutien était fort ($BF_{10} > 30$) pour la condition *enquête attributive*.

Tableau 16

Proportions de sélections de l'agent comme cause principale en fonction de ses motivations et du cadrage conversationnel

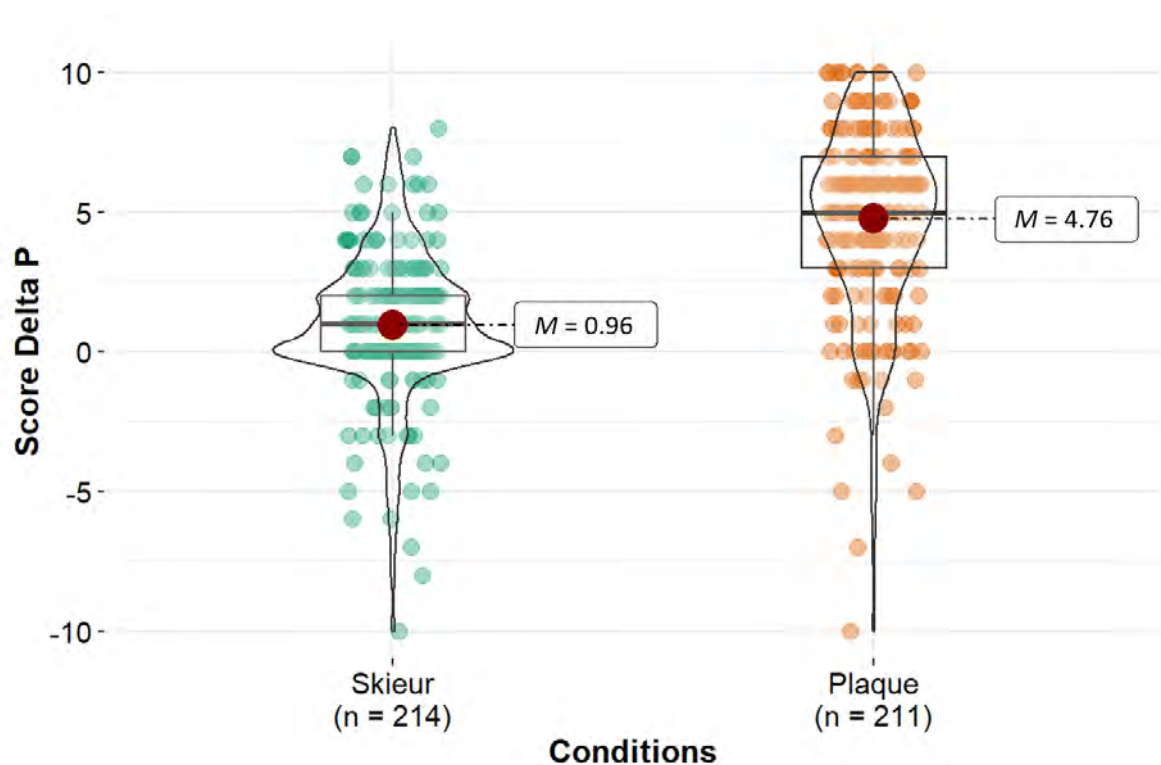
		Motivations		OR	p	BF ₁₀
		Louables	Blâmables			
Enquête	Explicative ($n = 130$)	.057	.200	4.13 [1.35, 15.47]	.020	2.87
	Attributive ($n = 130$)	.075	.302	5.35 [1.98, 17.14]	.002	46.22

Scores ΔP des conditions. Comme indiqué dans la Figure 16 (p.167), la plaque de glace a reçu un score ΔP plus élevé que le skieur, $t(210) = 13.9$, $p < .001$, $d = 0.96$, ICd 95% = [0.79, 1.12], $BF_{10} = 2.14 \times 10^{28}$. Le d de Cohen indiquait une large taille d'effet et le facteur

bayésien un soutien décisif pour l'hypothèse alternative. Une ANOVA 2 (Cadrage conversationnel : enquête explicative, enquête attributive) \times 2 (Motivations : louables, blâmables) a été conduite sur le score ΔP du skieur. Le score ΔP du skieur ayant des motivations blâmables ($M = 1.38$, $Mdn = 1$, $SD = 2.67$) était significativement plus élevé que celui du skieur ayant des motivations louables ($M = 0.56$, $Mdn = 1$, $SD = 2.46$), $F(1, 210) = 6.14$, $p = .014$, $\eta_p^2 = .028$. En revanche, les analyses n'ont pas montré d'effet principal du cadrage conversationnel sur le score ΔP , $F(1, 210) = 1.33$, $p = .250$, $\eta_p^2 = .006$. Nous n'avons pas non plus observé d'effet d'interaction statistiquement significatif entre ces deux prédicteurs, $F(1, 210) = 2.24$, $p = .136$, $\eta_p^2 = .011$.

Figure 16

Distribution des scores Delta P du skieur et de la plaque de neige.



Discussion de l'étude 6

L'objectif de cette étude était de confronter le CCM et l'approche pragmatique en proposant une opérationnalisation plus fidèle du cadrage conversationnel, notamment de la condition *enquête attributive*. Contrairement à l'Étude 5 qui n'a pas répliqué l'effet de la valence morale sur la sélection causale, cet effet a été répliqué dans l'Étude 6, aussi bien dans la condition *enquête attributive* que dans la condition *enquête explicative*. Il s'agissait de fortes tailles d'effet dans les deux conditions. Néanmoins, les facteurs bayésiens apportaient considérablement plus de soutien à l'hypothèse de l'existence d'un effet de la valence morale dans l'*enquête attributive* (soutien fort) que dans l'*enquête explicative* (soutien anecdotique). Ces résultats pourraient aller dans le sens d'une atténuation de l'effet de la valence morale avec le cadrage conversationnel. Même si ces résultats pourraient apporter un soutien à l'approche pragmatique du jugement causal, l'hypothèse était que l'effet de la valence morale serait neutralisé dans la condition de l'*enquête explicative*. Par conséquent, l'ambiguïté conversationnelle pourrait éventuellement expliquer une part de cet effet, mais pas la totalité.

Par ailleurs, le résultat selon lequel le score ΔP de l'agent blâmable était plus élevé que celui de l'agent louable, et ce, indépendamment du cadrage conversationnel, apporte un soutien au CCM. Ce score ayant été conçu pour référer sans ambiguïté à la causalité au sens strict (Cheng & Novick, 1992 ; Samland & Waldmann, 2016), ce résultat indiquerait que la représentation du lien causal entre l'agent et le dommage serait bel et bien biaisée par une volonté de blâmer l'agent. Cela expliquerait pourquoi l'effet de la valence morale est également retrouvé avec la question ouverte, et ce, quel que soit le cadrage conversationnel de la tâche (i.e., *enquête explicative* ou *enquête attributive*). Les motivations de l'agent ont également eu un effet sur la sélection de l'agent comme élément qui a le plus contribué au dommage. Étant donné qu'il n'y a pas eu d'interaction avec le cadrage conversationnel, nous ne pouvons pas considérer que cette question ait été interprétée comme une *enquête attributive*. De plus, ni la

question du contributeur le plus important, ni la question ΔP ne comprenaient le mot *cause* -- terme jugé ambigu par les tenants de l'approche pragmatique de l'effet de la valence morale (Samland & Waldmann, 2016). Enfin, l'hypothèse pragmatique selon laquelle les questions centrées sur l'agent seraient celles qui inciteraient les participants à interpréter la tâche comme une *enquête attributive* n'est pas non plus soutenue par nos résultats. Spécifiquement, les motivations ont eu un effet sur le jugement causal dans cette étude tant pour les questions ouvertes, telles que celles sur la cause principale ou l'élément le plus contributeur, que pour la question à échelle centrée sur l'agent avec la mesure ΔP .

Les résultats de cette étude convergent ainsi vers le soutien du CCM plutôt que de l'approche pragmatique. Cependant, ce soutien est à nuancer car l'effet de la valence morale ne semble concerner qu'une minorité des participants. En effet, même dans la condition la plus favorable au *processus de validation du blâme* (i.e., motivations blâmables et enquête attributive), seuls 30% des participants ont désigné l'agent comme cause principale du décès des habitants. Conformément au critère normatif de sensibilité, une large majorité des participants a désigné la plaque de glace comme cause principale et comme élément ayant le plus contribué au décès des habitants. Les participants lui ont également attribué le score ΔP le plus élevé.

Discussion du Chapitre 4

L'objectif du présent chapitre était d'apporter des éléments de réponse à la question fondamentale de savoir si l'effet de la valence morale sur le jugement causal peut bel et bien refléter un biais ou, au contraire, une ambiguïté dans la tâche de jugement causal (Fischhoff, 1982 ; Willemsen & Kirfel, 2019). Alors que le CCM soutient la première possibilité (i.e., les individus sont victimes d'un *processus de validation du blâme*), l'approche pragmatique soutient la seconde (i.e., le mot *cause* est polysémique).

Afin d'apporter des éléments de réponse à ce débat, nous avons présenté une revue narrative de la littérature, complétée de deux études en adoptant une perspective pragmatique de l'effet de la valence morale. A travers la revue narrative, nous avons étendu l'hypothèse de Samland et Waldmann sur la focalisation de la mesure du jugement causal, en nous intéressant aux scénarios présentés et notamment aux différents modèles causaux de ces scénarios. Nous avons proposé que la totalité du matériel employé dans la littérature incite les participants à interpréter la tâche comme une *enquête attributive*. Ce constat nous a amené à proposer une méthodologie plus neutre (i.e., ne favorisant pas une interprétation particulière de la causalité) que celles utilisées majoritairement dans les précédentes recherches, à savoir : un modèle causal possédant une bonne réponse aux deux types d'enquêtes (explicative/attributive) et une question ouverte (non focalisée sur l'agent). A partir de cette base présumée neutre d'un point de vue interprétatif, nous avons manipulé le cadrage conversationnel de la question posée au sein des Études 5 et 6, favorisant ainsi chez les participants une interprétation de la question causale soit en termes de recherche d'explication (*enquête explicative*), soit en termes de recherche de responsable de l'incident (*enquête attributive*). L'hypothèse pragmatique était que l'effet de la valence morale serait répliqué lorsque la tâche est cadrée comme une *enquête attributive*, et neutralisé lorsque la tâche est cadrée comme une *enquête explicative*.

En premier lieu, il convient de noter que les résultats des études 5 et 6 ont montré que la cause principale majoritairement désignée par les participants était systématiquement la condition nécessaire réelle la moins sensible aux circonstances (i.e., la plaque de neige). Ceci vient confirmer l'ensemble des résultats obtenus dans cette thèse, à savoir que les jugements causaux de sens commun sont principalement influencés par les critères normatifs de la causalité, à savoir la nécessité réelle et la sensibilité. Cette préférence pour la condition physique contredit certains résultats de la littérature ayant montré que les actions humaines étaient davantage sélectionnées que les événements physiques (Hilton et al., 2016 ; McClure et

al., 2007). Néanmoins, ces auteurs ont utilisé un modèle causal que nous avons nommé *Modèle causal 2* (i.e., aucune condition nécessaire ne se distinguait par sa force causale). Il n'y avait donc pas de raison apparente de sélectionner rationnellement la condition physique comme cause principale. A l'inverse, dans notre scénario, la condition physique (i.e., la plaque de neige) possédait la plus grande force causale et a effectivement reçu le score ΔP le plus élevé. Par conséquent, cette divergence entre nos résultats et ceux observés dans de précédentes études semblerait relativement cohérente : les individus sélectionneraient l'agent humain lorsqu'il n'y a pas de raison rationnelle de sélectionner une autre condition.

La sélection de l'agent humain comme cause principale n'a été faite que par une minorité de participants dans nos deux études, et ce, même dans des conditions supposées très favorables au *processus de validation du blâme* (i.e., motivations blâmables et enquête attributive). En d'autres termes, une minorité d'individus présenterait un jugement causal biaisé. Dès lors, il pourrait être intéressant d'étudier les caractéristiques de ces participants et de comprendre pourquoi leurs jugements s'éloignent des jugements rationnels fournis par la majorité des participants. En effet, plusieurs travaux récents ont souligné l'importance de prendre en compte les différences individuelles dans la recherche sur les biais cognitifs (Berthet, 2021 ; De Neys & Bonnefon, 2013 ; Frey et al., 2018 ; Thompson et al., 2018 ; Thompson & Markovits, 2021). Stanovich (2018) a, par exemple, proposé qu'un biais pouvait être dû à un manque de connaissances des critères normatifs, une incapacité à détecter qu'un processus entre en conflit avec l'application des critères normatifs, ou une incapacité à inhiber la réponse biaisée. Si nous appliquons cette classification à l'effet de la valence morale dans nos deux dernières études, il se pourrait que la minorité de participants qui mentionnent l'agent comme cause principale n'aient pas intégré le critère de sensibilité, ne détectent pas que la volonté de blâmer l'agent entre en conflit avec l'application de ce critère normatif, ou qu'ils ne

parviennent pas à inhiber le *processus de validation du blâme*. De futures études pourraient explorer ces questions plus en détails.

L'hypothèse pragmatique, selon laquelle l'effet de la valence morale serait lié à une interprétation de la tâche comme une *enquête attributive*, n'a pas été corroborée. A travers l'Étude 5, l'effet de la valence morale n'a pas été répliqué pour la question ouverte, cela quel que soit le cadrage conversationnel de la question causale. Ces résultats ne permettent de soutenir ni les arguments défendus par les tenants du CCM, ni ceux défendus par les tenants de l'approche pragmatique. Néanmoins, le score ΔP de l'agent, utilisé par Samland et Waldmann (2016) comme une mesure non-ambiguë du jugement de causalité au sens strict, était influencé par la nature de ses motivations (louables, blâmables). En plus de retrouver cet effet des motivations de l'agent sur le score ΔP , les résultats de l'Étude 6 ont également mis en avant un effet des motivations sur la sélection de l'agent comme cause principale et comme élément qui a le plus contribué au dommage. Ces effets observés sur le score ΔP dans nos deux études contredisent donc les résultats observés par Samland et Waldmann (2016). Comme évoqué dans la discussion de l'Étude 5, cette divergence pourrait s'expliquer par un dommage plus sévère dans nos études. En effet, les scénarios de Samland et Waldmann présentaient des dommages de faible gravité (i.e., une pénurie de stylos et une panne d'ordinateur), contrairement à notre scénario dans lequel plusieurs habitants perdirent la vie. L'évaluation spontanée négative envers notre agent pourraient ainsi avoir été plus importante que dans les études des deux auteurs. Conformément à l'hypothèse du CCM (Alicke, 2011), l'effet de la valence morale sur le jugement causal serait favorisé par une évaluation spontanée négative importante.

Contrairement à nos hypothèses, nous n'avons observé aucun effet d'interaction entre les motivations de l'agent et le cadrage conversationnel dans les études 5 et 6. Cela tendrait

donc à soutenir les hypothèses du CCM, alors même que la méthode utilisée a été conçue pour éviter les variables parasites aux inférences conversationnelles de nos participants (i.e., modèle causal 3 et question ne mentionnant pas l'agent). Par ailleurs, étant donné que l'effet de la valence morale a été répliqué avec des mesures n'incluant pas le mot *cause* (i.e., mesure ΔP et facteur ayant le plus contribué), ces résultats indiqueraient que, contrairement à l'hypothèse de Samland et Waldmann, l'ambiguïté du terme *cause* n'est pas responsable de l'effet de la valence morale.

En somme, nos résultats démontrent que, avec une méthodologie plus neutre (modèle 3 + question ouverte), la grande majorité des participants fournit une réponse rationnelle en désignant la condition la moins sensible aux circonstances (i.e., la plaque de glace). Ces résultats permettent d'étendre encore davantage nos conclusions sur l'interaction entre les critères normatifs et la valence morale dans les jugements causaux de sens commun. Néanmoins, une minorité de participants sélectionne l'agent et semble vulnérable à la valence morale de ses états mentaux. De plus, leur jugement ne semble pas, ou peu, affecté par le cadrage conversationnel de la tâche. Cela soutiendrait l'existence d'un jugement causal intrinsèquement biaisé par la valence morale de l'agent pour cette minorité de participants (Alicke, 1992, 2000 ; Alicke et al., 2008, 2011 ; Rogers et al., 2019). D'un point de vue applicatif, il apparaît nécessaire d'étudier les caractéristiques de ces participants. De plus, des méthodes de débiaisement pourraient être plus appropriées qu'une amélioration de la communication ou de la présentation des informations pour éliminer l'effet de la valence morale.

Ces études présentaient cependant quelques limites qu'il convient de présenter et de discuter pour nuancer nos conclusions. L'opérationnalisation du cadrage de la question peut notamment être interrogée. D'une part, la validité écologique de l'utilisation d'interlocuteurs

fictifs peut être discutée. En effet, leur présence fictive n'était peut-être pas assez prégnante pour donner suffisamment de crédibilité au cadrage conversationnel. Certains travaux ont en effet montré que la présence réelle d'autrui a un impact fort sur les jugements (Lerner & Tetlock, 1999). Dans cette perspective, l'interprétation de la question par les participants en termes d'enquête explicative ou attributive pourrait ne pas avoir été affectée par la variable *cadrage conversationnel*. Une autre potentielle limite à ces études concernerait le cadrage sur l'*enquête explicative*. En effet, même si le travail d'un journaliste est de relayer des informations pour faire comprendre un événement à une population, il n'est pas exclu que cette fonction active également des représentations en lien avec l'attribution de responsabilité. En effet, de nombreuses affaires judiciaires sont aujourd'hui relayées par les médias. Ces mêmes médias sont parfois critiqués pour leur manque de neutralité et il leur est reproché d'être source d'influence sur les décisions judiciaires (e.g., Arnaud & Ouss, 2016 ; Charon, 2008). Nous ne pouvons ainsi totalement exclure que le journalisme et les enquêtes attributives soient associés dans les représentations en mémoire des participants. Si tel est le cas, il est alors possible de considérer que les participants qui ont désigné l'agent comme cause principale dans le cadre de l'enquête explicative aient, en réalité, interprété la question en termes d'enquête attributive. Pour conclure, ces limites écologiques et de validité des construits souligneraient la pertinence de proposer de meilleures opérationnalisations des enquêtes explicatives et attributives. Cette amélioration paraît nécessaire avant de rejeter totalement une approche aussi récente que la perspective pragmatique de l'effet de la valence morale.

Discussion générale

« Here the epicenter of controversy shifts from the empirical robustness of effects to the normative benchmarks for classifying effects as erroneous. Following John Milton in Paradise Lost, we divide the contending theorists into two camps: the traditionalists, who seek to explain the ways of God to humans (by upholding traditional normative standards), and the revisionists, who seek to explain the ways of humans to God (by defending the reasonableness of human behavior). »

Tetlock et Mellers (2002, p. 98)

Ce travail de thèse avait pour objectif d'étudier la rationalité des jugements causaux dans les situations à l'issue négative impliquant un agent humain. Plus spécifiquement, il s'agissait de déterminer dans quelle mesure ce jugement est biaisé par la valence morale de l'agent. Pour répondre à cette question, nous avons mené une série d'études divisée en deux axes de recherche complémentaires. Le premier axe (Études 1 à 4) visait à étudier l'articulation entre l'effet de la valence morale et des critères normatifs philosophiques et juridiques de la causalité (i.e., nécessité réelle et sensibilité). Le deuxième axe (Revue narrative + Études 5 et 6), quant à lui, visait à examiner l'articulation entre cet effet de la valence morale et des critères normatifs conversationnels (i.e., maximes de relation et de quantité). En d'autres termes, il s'agissait de déterminer si l'effet de la valence morale pouvait être expliqué par le respect de ces critères conversationnels par les individus, plutôt que par un biais de jugement causal. L'objectif de ce dernier Chapitre est de dresser le bilan des études réalisées dans ces deux axes, leurs limites ainsi que les perspectives de recherche à envisager.

DISCUSSION GÉNÉRALE

Dans une première partie, nous procéderons à une synthèse des contributions apportées par ce travail. Nous commencerons en détaillant la manière dont les individus semblent connaître et appliquer prioritairement les critères normatifs de la causalité. Ensuite, nous discuterons des résultats ayant permis d'éclairer l'interaction entre l'application de ces critères et l'effet de valence morale. Par la suite, nous détaillerons les résultats permettant d'apporter des éléments de réponse à la question de savoir si l'effet est lié à un biais de jugement ou au respect de la logique conversationnelle. Enfin, nous aborderons l'importance de considérer de potentielles différences inter-individuelles dans la rationalité des jugements causaux, en soulignant les questionnements à cet égard que suscitent les résultats de nos études.

Dans une seconde partie, nous discuterons des limites des études que nous avons menées, ainsi que les nombreuses perspectives qu'elles ouvrent. Nous discuterons des limites liées à la validité écologiques de nos études, à la définition de la rationalité adoptée dans cette thèse, à la validité des construits (i.e., capacité à opérationnaliser les concepts théoriques étudiés) et à la validité externe (i.e., capacité à généraliser nos résultats à l'ensemble de la population humaine). Enfin, nous conclurons notre travail en soulignant les principaux enseignements tirés de nos recherches.

Apports de la thèse

Au regard de précédents travaux menés spécifiquement sur les critères normatifs de nécessité réelle et de sensibilité, nous nous attendions à ce que les individus connaissent et appliquent ces critères pour formuler leurs jugements de causalité. Plus précisément, nous nous attendions à ce que l'agent soit sélectionné comme la cause d'un dommage (tâche de jugement sous la forme de question ouverte) ou reçoive un score de causalité non nul (sous la forme d'une question fermée à échelle de réponse), uniquement si ses actions constituaient une condition *nécessaire réelle* au dommage. De plus, nous nous attendions à ce que celui-ci soit davantage sélectionné, et reçoive un score de causalité plus élevé, lorsque ses actions étaient

DISCUSSION GÉNÉRALE

insensibles aux circonstances de fond de l'événement (vs. sensibles). Enfin, cet axe visait à déterminer si une forte volonté de blâmer un agent en raison de sa valence morale irait jusqu'à inciter les individus à le désigner comme cause, et ce, alors même que celui-ci n'est pas une condition nécessaire réelle au dommage. Ceci nous permettait également de tester l'hypothèse du CCM selon laquelle l'effet de la valence morale se produirait uniquement lorsque le rôle causal de l'agent était ambigu, à savoir dans le cas d'un lien causal faible (vs. lien causal fort ou absent). En d'autres termes, l'effet de la valence morale sur le jugement de causalité devrait être neutralisé lorsque l'agent était clairement (i.e., lien causal fort) ou n'était clairement pas (i.e., lien causal absent) un contributeur important au dommage.

Connaissance et application prioritaire des critères normatifs de la causalité

Si certains travaux ont montré que les participants respectaient les critères de nécessité réelle et de sensibilité vis-à-vis d'événements physiques (Gerstenberg et al., 2017 ; Grinfeld et al., 2020) ou avec un dommage de faible gravité (Hilton et al., 2021), nos résultats ont permis d'étendre ces conclusions aux événements propices au *processus de validation du blâme*, à savoir : (1) lorsqu'un agent humain a des motivations blâmables (e.g., récupérer l'argent d'une victime, allumer un feu criminel) et (2) qu'un dommage sévère s'est produit (i.e., homicide, incendie meurtrier, empoisonnement).

En effet, les résultats des Études 1 à 4 convergent vers la connaissance et l'application du critère de nécessité réelle par les participants. L'Étude 1 a montré que l'agent n'est pas, ou peu (1 participant sur 140), désigné comme cause principale du dommage par les participants lorsque ses actions ne constituent pas une condition nécessaire réelle à celui-ci (i.e., lien causal préempté). En optant pour une question causale centrée sur l'agent à échelle de mesure, les Études 2 à 4 ont étendu ce résultat aux mesures permettant d'évaluer plus finement l'évaluation de la contribution causale de l'agent au dommage. Les résultats de ces études ont ici montré

DISCUSSION GÉNÉRALE

que le score de causalité attribué à l'agent était considérablement réduit dans le modèle causal préempté. L'hypothèse concernant le critère de nécessité réelle est donc en grande partie corroborée à travers ces 4 études, et permet d'apporter un premier élément de réponse à la question de la rationalité des jugements causaux dans les situations favorables au *processus de validation du blâme*. Il reste cependant à éclairer une observation inconsistante avec notre hypothèse, à savoir : les scores de causalité médians et moyens n'étaient pas systématiquement nuls dans les conditions préemptées. L'étude des distributions des scores de jugement causal pour cette condition a montré qu'une majorité de participants attribuait rationnellement le score de 1 (= pas du tout la cause) et une minorité lui attribuait un score non nul. Le fait qu'une minorité de nos participants ait pu sélectionner l'agent comme la cause d'un dommage, alors même que ses actions sont sans lien avec celui-ci, interroge de possible facteurs inter-individuels dont nous discuterons à la fin de cette partie.

A travers les Études 1, 2, 5 et 6, nous avons également mis en évidence que les participants connaissent et appliquent le second critère normatif de la causalité, à savoir le critère de sensibilité. En effet, l'agent était moins désigné comme cause principale (Étude 1) et son score de causalité était considérablement réduit (Étude 2) lorsque ses actions étaient sensibles (vs. insensibles) aux circonstances de fond. Par ailleurs, les participants préféraient majoritairement désigner comme cause principale et attribuer le score le plus élevé à la condition la moins sensible aux circonstances (i.e., la plaque de glace, Études 5 et 6).

Interaction entre l'application des critères normatifs et l'effet de la valence morale

Concernant la relation entre les deux critères normatifs et l'effet de la valence morale, les résultats de nos études sont également consistants sur l'application prioritaire de ces critères. L'effet de la valence morale sur le jugement causal n'a jamais été observé dans nos modèles causaux préemptés, c'est-à-dire lorsque l'agent n'était pas une condition nécessaire réelle au

DISCUSSION GÉNÉRALE

dommage (Études 1 à 4). En effet, nos travaux ont démontré que les participants connaissent et appliquent ce critère de façon prioritaire et ce, même lorsqu'un agent blâmable évalué de façon très négative (cf. les très fortes tailles d'effet des motivations de l'agent sur les scores d'évaluation dans toutes nos études) est impliqué dans un dommage légal. En d'autres termes, nous avons montré que le critère de nécessité réelle neutralise l'effet de la valence morale dans la situation pourtant la plus optimale pour l'émergence du *processus de validation du blâme* : un agent aux motivations blâmables souhaitait produire un dommage légal (i.e., meurtre, incendie criminel et empoisonnement alimentaire), il a agi pour provoquer le dommage et le dommage est survenu.

Si l'effet de la valence morale n'a pas été répliqué avec un lien causal préempté, il l'a cependant été avec un lien causal fort unissant les actions de l'agent au dommage (Études 1 et 2). Par conséquent, ceci vient partiellement corroborer l'hypothèse du CCM (Alicke et al., 2011), selon laquelle lorsqu'un agent est évidemment ou évidemment pas un contributeur important à l'événement, l'effet de la valence morale n'apparaîtrait pas. Nos résultats ont montré que lorsque l'agent était un contributeur important (i.e., lien causal fort), l'effet de la valence morale était répliqué avec une large taille d'effet, aussi bien avec une question ouverte (Étude 1, $OR = 8.85$) qu'avec une question à échelle de mesure (Étude 2, $d = 0.93$). Cela implique que l'effet de la valence morale serait neutralisé, seulement s'il n'y a pas de lien causal réel entre l'agent et le dommage. Contrairement à l'hypothèse du CCM, le critère neutralisant l'effet ne serait pas l'ambiguïté du lien causal, mais bien la possibilité de produire un jugement respectant le critère normatif de nécessité réelle. Ces résultats permettent ainsi d'enrichir le CCM et de soutenir l'hypothèse d'une application prioritaire du critère de nécessité réelle dans les situations les plus favorables au *processus de validation du blâme*.

DISCUSSION GÉNÉRALE

Par ailleurs, lorsqu'une question ouverte était posée (Étude 1), nous avons observé un effet de neutralisation de l'effet de la valence morale par la force du lien causal (fort vs. faible) à l'opposé des prédictions du CCM (Alicke et al., 2011). Formulé autrement, l'effet de la valence morale n'a été observé qu'en présence d'un lien causal fort, donc lorsque le lien causal n'était pas ambigu, et que l'agent était évidemment un contributeur important au dommage. Une explication pourrait être que lorsque le lien causal entre l'agent et le dommage était faible, et donc sensible à une autre condition (i.e., les ingrédients allergènes du nouveau plat), les participants ne disposaient pas de suffisamment d'arguments rationnels pour désigner l'agent comme cause principale du dommage. Cette interprétation pencherait également en faveur d'une rationalité des jugements causaux : les individus préféreraient désigner comme cause principale la condition ayant la plus grande force causale plutôt que le facteur le plus blâmable. En effet, dans une tâche de jugement causal mobilisant une question ouverte et un lien causal faible entre l'agent et le dommage, la grande majorité des participants sélectionnent la condition la moins sensible aux circonstances (i.e., allergie, plaque de neige) comme cause principale du dommage (76% dans l'Étude 1 ; 71% dans l'Étude 5 ; 72% dans l'Étude 6). Bien que les Études 2, 5 et 6 aient montré un effet de la valence morale sur les jugements causaux pour des situations où le lien entre l'agent et le dommage était faible, la condition recevant le score de causalité le plus élevé était également la condition la moins sensible (i.e., plaque de neige) lorsque les participants devaient attribuer un score à toutes les conditions (Études 5 et 6). Par conséquent, à l'instar du critère de nécessité réelle, le critère de sensibilité serait également appliqué de façon prioritaire par les individus par rapport à l'effet de la valence morale de l'agent. Ces résultats corroborent les travaux de Engelmann et Waldmann (2021, 2022), lesquels ont montré qu'un agent recevait un score de responsabilité plus élevé lorsqu'il était proche (vs. éloigné) du dommage dans la chaîne causale. Comme discuté dans l'Étude 1 (pp.83-84), ces résultats vont également à l'encontre de ceux de Hilton et ses collaborateurs

DISCUSSION GÉNÉRALE

(2010) ayant montré que les participants sélectionnaient l'agent blâmable comme cause principale bien qu'il soit éloigné dans la chaîne causale. Néanmoins, leurs récits présentaient un agent souhaitant provoquer le dommage. Cette distinction pourrait être importante car, selon Heider (1958), lorsqu'un agent souhaite provoquer un dommage, cet agent peut être considéré comme moins sensible aux circonstances puisque son objectif l'incitera à s'adapter à tout changement dans les circonstances. Il pourrait alors y avoir une confusion potentielle entre valence morale et force causale. Hilton et ses collègues (2010) ont effectivement observé que l'agent avait un score de force causale plus important lorsqu'il souhaitait (vs. ne souhaitait pas) provoquer le dommage. En maintenant constante (Étude 1), ou en neutralisant (Études 5 et 6), l'intention de l'agent de provoquer le dommage dans nos études, nous avons observé que les participants ne remontent pas la chaîne causale pour sélectionner l'agent comme cause principale. L'ensemble de ces résultats semble ainsi soutenir la prise en compte de la force causale de chaque condition de l'événement par les participants.

Effet de la valence morale : biais de jugement ou logique conversationnelle ?

Le deuxième axe de recherche de cette thèse visait à aller encore plus loin en testant l'hypothèse d'une parfaite rationalité des individus, à savoir que ceux-ci appliqueraient non seulement les critères normatifs de la causalité (Icard et al., 2017 ; Lewis, 1973 ; Woodward, 2006), mais également ceux de la communication humaine (Grice, 1975 ; Sperber & Wilson, 1996). Dans cette perspective, l'effet de la valence morale serait lié à l'ambiguïté du mot *cause* et à celle des tâches de jugement causal, renvoyant aussi bien à une *enquête explicative* (i.e., évaluation de la causalité au sens strict) qu'à une *enquête attributive* (i.e., recherche d'un responsable). Plus précisément, l'effet de la valence morale se manifesterait uniquement lorsque la question peut être interprétée comme une *enquête attributive*. L'erreur de jugement, identifiée dans de nombreuses études, résiderait alors dans la non-prise en compte de la logique conversationnelle par les expérimentateurs, plutôt que dans les processus cognitifs biaisés des

DISCUSSION GÉNÉRALE

participants. Les résultats des deux études de ce second axe n'ont pas permis de confirmer l'hypothèse selon laquelle le cadrage conversationnel neutraliserait ou atténuerait l'effet de la valence morale. En effet, aucune de nos deux études (Études 5 et 6) n'a permis d'observer d'effet d'interaction entre le cadrage conversationnel et l'effet de la valence morale. De plus, nous avons retrouvé l'effet de la valence morale avec une question ouverte mobilisant le mot *cause* (Étude 6), une mesure ΔP (Études 5 et 6), et une question sur le contributeur le plus important (Étude 6). Ces résultats contredisent l'hypothèse de Samland et Waldmann (2016) qui suggérait que l'ambiguïté de la tâche résidait dans le mot *cause*, car nous avons observé l'effet de la valence morale indépendamment de la formulation de la question. Par conséquent, il semblerait que l'effet de la valence morale corresponde à un biais de jugement des individus provoqué par un raisonnement motivé, plutôt qu'à une ambiguïté conversationnelle liée à la tâche et non prise en compte par l'expérimentateur.

Des différences inter-individuelles dans la rationalité des jugements causaux ?

Comme nous l'avons mentionné en début de discussion, il convient cependant de prendre en considération la minorité de participants produisant, malgré tout, des jugements causaux irrationnels vis-à-vis du critère de nécessité réelle. Concernant les études ayant mobilisé une question ouverte, cette minorité de participants concernait pour rappel les participants ayant sélectionné l'agent comme cause principale du dommage dans un modèle causal préempté (1 participant sur 140, Étude 1) et dans un modèle où l'agent ne possède pas le lien causal le plus fort (28 participants sur 147, Étude 5 ; 40 participants sur 260, Étude 6). Par ailleurs, lorsque des questions à échelle de mesure étaient mobilisées (Études 2, 3 et 4), une minorité de participants attribuait également un score non nul à l'agent lorsque ses actions ne constituaient pas une condition nécessaire réelle à la survenue du dommage (modèle causal préempté). Ceci soulève la question des différences interindividuelles dans la rationalité des jugements (Berthet, 2021 ; De Neys & Bonnefon, 2013 ; Frey et al., 2018 ; Thompson et al.,

DISCUSSION GÉNÉRALE

2018 ; Thompson & Markovits, 2021). En effet, il est possible qu'une partie des participants basent leur jugement uniquement sur les critères normatifs du jugement causal, notamment la majorité qui respectent scrupuleusement les deux critères normatifs. Une autre partie serait en revanche plutôt motivée à sélectionner l'agent comme cause principale du dommage, sélection pouvant résulter d'une volonté de blâmer comme soutenu par le CCM.

Ces différences inter-individuelles peuvent toutefois avoir des origines très variées (De Neys & Bonnefon, 2013). Premièrement, elles pourraient être liées à une intégration en mémoire plus ou moins importante des critères normatifs entre les individus (Stanovich, 2018). En effet, il est possible que la minorité que nous avons évoquée n'ait pas intégré les critères de nécessité réelle et/ou de sensibilité. Deuxièmement, il semblerait que le respect du critère de nécessité réelle soit enfreint essentiellement lorsque les participants doivent évaluer le degré de contribution causale de l'agent (Études, 2, 3 et 4). Une explication à cela pourrait être que certains participants ont une définition différente et particulière de la causalité. Par exemple, ils pourraient avoir une conception probabiliste de la causalité conformément au *Crediting Causality Model* de Spellman (1997) : les actions suffisantes de l'agent pour provoquer le dommage dans le modèle causal préempté augmente la probabilité que le dommage survienne à 100% (i.e., à partir du moment où Barbara verse le poison létal dans le verre, le décès de son mari est garanti). Si cette conception constitue une infraction au principe de causalité réelle, il ne peut être exclu qu'elle soit responsable des scores attribués par la minorité de participants dans le lien causal préempté. Troisièmement, ces différences entre les participants pourraient s'expliquer par des divergences dans les capacités d'inhibition (Frederick, 2005 ; Houdé, 2019), de détection du conflit (interférence) entre l'évaluation spontanée et l'évaluation objective des liens causaux (De Neys & Franssens, 2009 ; Frey et al., 2018), de régulation émotionnelle (Gross et al., 2006), ou plus généralement de compétences métacognitives (Fletcher & Carruthers, 2012 ; Pennequin, 2021). Quatrièmement, l'identification à l'agent ou

DISCUSSION GÉNÉRALE

à la victime pourrait moduler la volonté de blâmer l'agent (Burger, 1981) et, ainsi, modérer l'effet de la valence morale. Enfin, il serait également pertinent d'examiner le lien entre l'effet de la valence morale et Locus de contrôle des participants (LOC), à savoir le degré auquel l'individu croit qu'un événement s'explique par des facteurs internes ou externes (Rotter, 1954). En effet, de nombreuses études ont montré que les individus ayant un LOC interne (vs. LOC externe) se considèrent, non seulement eux-mêmes mais aussi autrui, comme plus responsable de leur situation (pour une revue, Bachrach et al., 1977). Nous pourrions alors imaginer que les individus ayant un LOC externe (vs. interne) ressentiraient une évaluation moins négative envers l'agent, car ils attribueraient son comportement à des facteurs externes à sa personne (e.g., son milieu social).

Résumé des apports de la thèse

En somme, nos résultats démontrent que les deux critères définissant un jugement causal idéal et rationnel (i.e., nécessité réelle et sensibilité) occupent une place centrale et dominante dans les jugements causaux de sens commun. La prise en compte de facteurs extra-normatifs, telle que la valence morale de l'agent témoignerait malgré tout d'une part d'irrationalité dans ces jugements. Cette part d'irrationalité semblerait, à ce jour, liée aux individus eux-mêmes, plutôt qu'à une interprétation de la tâche qui ne correspond pas aux attentes de l'expérimentateur. En considérant, comme d'autres auteurs (Brainerd & Reyna, 2001 ; Etzioni, 2014), la rationalité sur un continuum, nous pouvons prétendre que le biais de jugement est toutefois fortement restreint par l'application prioritaire des deux critères normatifs de la causalité. En effet, si le CCM considère que les évaluations spontanées biaisent l'évaluation objective des liens causaux, nos travaux ont permis de distinguer les limites de cette influence. L'effet de la valence morale constituerait un effet additionnel indésirable survenant majoritairement si les individus ont la possibilité de respecter les critères normatifs de nécessité réelle et de sensibilité. De plus, il semblerait que l'évaluation négative envers

DISCUSSION GÉNÉRALE

l'agent doit être suffisamment forte pour que l'effet de la valence morale apparaisse (cf. dommage légal vs. non-légal ; Études 3 et 4). Nos données viennent ainsi soutenir l'idée selon laquelle les jugements causaux de sens commun sont associés à une « force normative » (Waldmann, 2017), c'est-à-dire une volonté de comprendre le monde tel qu'il est afin de pouvoir s'y adapter efficacement. Pour ce faire, la majorité des individus a ainsi intégré des critères de jugement permettant de détecter la présence d'un lien causal (critère de nécessité réelle) et d'en mesurer la force (critère de sensibilité). Cette rationalité des jugements causaux pourrait également s'expliquer par une volonté de présenter des arguments suffisamment rationnels pour convaincre autrui (Mercier & Sperber, 2011).

D'un point de vue appliqué, ces résultats nous renseignent sur la manière de débiaiser les jugements causaux. Ils indiqueraient notamment que les personnes amenées à juger une situation n'iraient pas jusqu'à inventer un lien causal inexistant pour blâmer un accusé suscitant une évaluation négative. Cette conclusion est soutenue par les résultats de nos Études 3 et 4, qui ont également montré que l'effet de la valence morale sur l'attribution de responsabilité était neutralisé par l'absence d'un lien causal. Par conséquent, la présence d'un lien causal requise par la majorité des systèmes juridiques pour établir une culpabilité serait donc respectée par les jugements causaux de sens commun. De plus, il semblerait également que les jugements des individus ne remonteraient pas particulièrement loin dans les chaînes causales pour désigner un agent coupable d'un dommage, respectant ainsi le critère de force causale (i.e., sensibilité). En effet, au vu des résultats des Études 1, 2, 5 et 6, nous pouvons imaginer que le facteur sélectionné ne doit pas être trop sensible aux circonstances pour être désigné comme cause, et ce, même si un agent blâmable se trouve en amont de la chaîne causale. Ceci soutient la théorie normative de la causalité adéquate prescrite par le droit français (cf. pp. 25-27). Par ailleurs, nos travaux permettent de guider de futures études sur le débiaisement du jugement causal. Premièrement, l'inefficacité du cadrage conversationnel sur l'effet de la valence morale

DISCUSSION GÉNÉRALE

suggérerait que la cible de l'intervention devrait être les individus plutôt que la tâche. Deuxièmement, le biais de jugement causal que constitue l'effet de la valence morale ne serait pas dû à la non-intégration en mémoire des critères normatifs, ou à une incapacité d'une majorité d'individus à les appliquer aux tâches qui nous intéressent. Par conséquent, il ne semble pas primordial que les méthodes de débiaisement intègrent une formation à ces critères et à leur application. Les futurs travaux devraient notamment tester des méthodes d'inhibition du *processus de validation du blâme* (Houdé et al., 2000 ; Moutier et al., 2002), ou de détection de l'interférence entre ce processus et l'évaluation objectif des liens causaux (De Neys, 2014). Ces travaux devront également porter une attention particulière aux différences inter-individuelles, telles que l'intégration en mémoire plus ou moins importante des critères de nécessité réelle et de sensibilité, la régulation émotionnelle ou encore l'identification ou les préjugés envers l'agent et la victime.

Limites et perspectives

Dans cette partie, nous pointerons des limites du travail réalisé tant sur le plan théorique que méthodologique, et décrirons des perspectives de recherches futures.

Si nous reprenons les critères de Wagner-Egger (2011) pour mettre au jour un véritable biais cognitif, le chercheur doit montrer que : (1) Il s'agit d'un processus systématique, (2) qui a des conséquences dans la vie quotidienne, (3) qui n'est pas lié à la communication avec l'expérimentateur, (4) que la tâche est écologiquement valide, (5) que la réponse jugée fautive n'est pas explicable par une rationalité concurrente et (6) que la réponse correcte est incontestée. L'effet de la valence morale sur le jugement causal est systématique et va toujours dans le même sens (1). Celui-ci présente des conséquences néfastes dans la vie quotidienne, notamment dans les décisions judiciaires (2). Les études 5 et 6 du présent travail ont montré

DISCUSSION GÉNÉRALE

que l'effet de la valence morale résistait au cadrage conversationnel de la question de causalité (3). Néanmoins, nous allons voir que les points (4), (5) et (6) peuvent être discutés.

La validité écologique des études sur l'effet de la valence morale

Une première limite générale, commune à l'ensemble des études empiriques sur l'effet de la valence morale, est effectivement celle de la validité écologique (4). Tester l'effet de la valence morale avec une tâche plus proche de contextes naturels pourrait être d'autant plus pertinent que certains « biais » ont été neutralisés lorsque les informations étaient présentées de façon plus écologique (e.g., Sedlmeier & Hilton, 2012 ; Shafir & LeBoeuf, 2002). Nous allons passer en revue quatre limites concernant la validité écologique : l'utilisation de situations et d'interlocuteurs fictifs, les passations individuelles, les termes utilisés dans les tâches et la distinction entre situations de certitude et situations d'incertitude.

Face à des situations et des interlocuteurs fictifs, les participants pourraient ne pas être aussi impliqués dans la tâche de jugement causal, tant d'un point de vue cognitif, émotionnel que motivationnel, qu'ils le seraient face à des situations réelles. Ce problème de validité écologique, commun à toutes les études sur l'effet de la valence morale, pourrait impacter le *processus de validation du blâme*, celui-ci étant modulé par l'intensité de la réaction affective à l'endroit de l'agent (Alicke et al., 2011). Par conséquent, un doute peut persister quant au respect du critère de nécessité réelle par les individus pour des situations dans lesquelles l'observateur éprouverait une évaluation négative intense envers un réel (vs. fictif) agent blâmable. Une hypothèse pourrait être que l'intensité des réactions affectives face à une situation réelle (vs. expérimentale) serait significativement plus forte et pourrait biaiser la représentation des individus au point d'inventer un lien causal inexistant. Par ailleurs, concernant l'utilisation d'interlocuteurs fictifs (journaliste, juge), nous pouvons nous demander envers qui les participants essaient d'être pertinents : l'interlocuteur fictif ou bien

DISCUSSION GÉNÉRALE

l'expérimentateur qui a mobilisé des interlocuteurs fictifs ? Comme le soulignait Hilton (1990), l'utilisation de questions à échelles de mesure, plutôt que de questions ouvertes, peut amener les participants à inférer qu'il s'agit davantage d'une « question d'examen » et qu'ils doivent montrer leurs connaissances à l'expérimentateur, qu'une conversation dans laquelle ils doivent être informatifs vis-à-vis des connaissances de leur interlocuteur. En développant cette analyse, nous considérons comme possible que, face à une étude annoncée comme une tâche de jugement ou de raisonnement, les participants considèrent les questions comme des « tests d'intelligence », et adaptent leurs réponses en anticipant une évaluation de la part de l'expérimentateur. Or, les travaux sur l'*accountability effect* (Lerner et al., 1998 ; Lerner & Tetlock, 1999) montrent effectivement que les biais de raisonnement sont réduits lorsque les individus anticipent une évaluation de la part d'autrui. Si les études expérimentales provoquent l'anticipation d'une évaluation de l'expérimentateur, il est possible que la rationalité des individus soit surestimée par rapport aux situations de la vie réelle dans lesquelles cette évaluation n'a pas toujours lieu. Conformément aux recommandations de Lerner et Tetlock (1999), une méthode pour contrôler cela serait d'intégrer des mesures de l'anxiété sociale, du self-monitoring ou de l'individuation dans les prochaines études. Par exemple, si l'effet de la valence morale est négativement corrélé à l'anxiété sociale, nous pouvons émettre l'hypothèse que les participants, les plus soucieux de l'image qu'ils renvoient, inhibent le *processus de validation du blâme* pour « plaire » à l'expérimentateur.

Une deuxième limite de validité écologique est celle de la passation individuelle des tâches de jugement causal. Or, selon la théorie argumentative du raisonnement (Mercier & Sperber, 2011), le raisonnement est une activité essentiellement sociale, notamment pour produire des arguments et convaincre nos semblables dans des discussions de groupe. De plus, les décisions judiciaires se prennent également en groupe à l'issue de discussions. Or, la recherche en psychologie sociale a établi depuis longtemps que les décisions en groupe

DISCUSSION GÉNÉRALE

présentent des caractéristiques différentes des décisions individuelles. La question de savoir si les décisions groupales amplifient ou atténuent les biais de jugement ne trouve pas de réponse claire, et pour cause : il y aurait de nombreux modérateurs en jeu (Bang & Frith, 2017 ; Kerr & Tindale, 2004).

Lorsque la tâche comporte une réponse correcte démontrable (e.g., un problème mathématique), il suffit qu'un des membres du groupe la trouve pour que l'ensemble du groupe soit convaincu par ses arguments (Laughlin & Ellis, 1986). Il serait intéressant de tester si l'application des critères normatifs du jugement causal (nécessité réelle et sensibilité) constitue une réponse correcte démontrable—par des arguments tels que « Si l'agent n'avait pas fait de hors-piste, alors la plaque de neige ne se serait pas décrochée » —et susceptible de convaincre l'ensemble du groupe dès lors qu'un membre l'exprime. Si c'est le cas, la prise de décision groupale devrait amplifier le respect des critères normatifs par les individus : la majorité des participants qui applique ces critères de façon exclusive pourrait convaincre la minorité, qui désigne l'agent blâmable lorsqu'il n'est pas la condition la plus sensible, de les appliquer également.

D'un autre côté, l'effet de polarisation (Sunstein, 2002) peut laisser supposer que la validation du blâme serait plus extrême en groupe. Cet effet renvoie aux prises de décisions plus extrêmes des groupes par rapport aux positions initiales de ces membres. Une étude de Myers et Kaplan (1976) a notamment montré que des groupes de jurés fictifs avaient des verdicts d'innocence ou de culpabilité plus extrêmes après avoir discuté en groupe. Néanmoins, un modérateur potentiel de l'effet de polarisation en groupe est la présence d'un consensus en son sein : l'effet est amplifié lorsque les avis des membres sont trop similaires (Bang & Frith, 2017). Nous pouvons alors imaginer que si un agent enfreint une norme morale qui fait consensus au sein du groupe (e.g., voler les affaires d'un blessé), un effet de polarisation

DISCUSSION GÉNÉRALE

pourrait amplifier l'effet de la valence morale. Néanmoins, lorsque les avis divergent⁵⁰ au sein du groupe, les décisions peuvent être améliorées (Kogan & Wallach, 1966 ; Moshman & Geil, 1998 ; Vinokur & Burnstein, 1978). Par conséquent, nous pourrions nous attendre à ce que l'effet de la valence morale soit atténué lorsque le jugement causal est effectué dans un groupe dont les membres ont des avis divergents. Cette interprétation va dans le sens des résultats de Monroe et Malle (2019) qui ont montré que les jugements de blâme étaient réduits lorsque les participants recevaient des informations au cours de la tâche qui atténuaient la négativité du comportement de l'agent. Il serait également intéressant de tester si le fait de demander à une partie des membres du groupe de se faire « l'avocat du diable », en défendant la position opposée au consensus du groupe, atténuerait l'effet de la valence morale (Herbert & Estes, 1977).

Une troisième limite liée à la validité écologique a été mise en lumière par Rose et ses collaborateurs (2021). Ces auteurs sont partis du constat que la quasi-totalité des études expérimentales sur le jugement causal mobilise le mot *cause* dans la question de causalité (e.g., « Dans quelle mesure l'agent est-il la cause de l'événement ? »). Or, le langage courant mobilise davantage des verbes causaux spéciaux pour parler de causalité, tels que *brûler*, *casser* ou *tuer*. Ces chercheurs ont notamment observé que le mot *casser* est, à lui seul, plus fréquent dans le langage courant que le mot *cause*. Ils ont voulu tester l'hypothèse selon laquelle le mot *cause* renvoie aux théories philosophiques de dépendance contrefactuelle, alors que les verbes causaux (e.g., *tuer*) renvoient davantage aux théories du processus (ces deux approches philosophiques ont été présentées dans le Chapitre 1, pp. 14-15). Les résultats obtenus ont partiellement corroboré leur hypothèse en montrant des différences dans les jugements causaux des participants selon le terme employé dans la question (verbes causaux spéciaux vs. « a

⁵⁰ Jusqu'à un certain point car une trop forte divergence peut mener au conflit et causer des problèmes de communication et de mauvaises prises de décision (Bang & Frith, 2017).

DISCUSSION GÉNÉRALE

causé »). Il pourrait alors être intéressant d'étudier l'effet de la valence morale et le respect des critères normatifs en utilisant des verbes causaux spéciaux comme Rose et ses collaborateurs (e.g., « Dans quelle mesure êtes-vous d'accord avec l'affirmation suivante : Barbara a tué son mari ? »).

Enfin, une dernière limite à considérer concerne la distinction entre les situations de certitude et les situations d'incertitude. Comme toutes celles conduites sur l'effet de la valence morale, nos études s'intéressent aux situations de certitudes, à savoir des situations où les participants doivent effectuer un jugement causal à partir d'informations sur l'événement présumées vraies. Or, dans le contexte d'une décision judiciaire, il est fréquent que la tâche de jugement débute dans une situation d'incertitude, notamment lorsque les versions des faits de l'accusé et de la victime sont contradictoires. Par exemple, l'accusé peut présenter une version des faits dans laquelle ses actions ne sont pas une condition nécessaire réelle au dommage, tandis que la victime soutient une version des faits où c'est effectivement le cas. Dans de telles situations d'incertitude, le modèle psychologique dominant est le *Story Model* (Hastie, 1993 ; Pennington & Hastie, 1986, 1992 ; Willmott et al., 2018). Selon ce modèle, les jurés utiliseraient à la fois les preuves présentées au cours du procès, leurs connaissances du monde et leurs attitudes pour construire une ou plusieurs interprétations possibles de l'événement, appelées « récits ». Une fois ces récits établis, les jurés évaluent la probabilité que chaque récit soit vrai en se basant sur les preuves présentées lors du procès, la cohérence interne du récit et sa plausibilité. Une fois le récit le plus probable identifié, les jurés évaluent la culpabilité de l'accusé en se basant notamment sur le modèle causal de ce récit.

Il serait intéressant d'examiner si la valence morale de l'agent a un impact sur l'évaluation de chaque récit. A partir du CCM, nous pourrions formuler l'hypothèse que, face à un agent possédant une valence morale négative, les individus tendent à évaluer la version de

DISCUSSION GÉNÉRALE

la victime comme étant plus probable que celle de l'accusé. A l'inverse, la version de la victime pourrait être discréditée si l'agent possède une valence morale positive. Cette étude offrirait l'opportunité d'explorer l'effet de la valence morale dans un contexte plus écologique vis-à-vis des décisions judiciaires, en examinant les jugements causaux en situations d'incertitude.

Plusieurs définitions de la rationalité

Concernant les points (5) et (6), nous pouvons également interroger la définition de la rationalité mobilisée dans cette thèse. Cette question a d'ailleurs été l'objet d'un « Grand débat sur la rationalité » en sciences cognitives (Tetlock & Meyers, 2002), opposant les chercheurs qui critiquent l'utilisation de la rationalité normative en psychologie expérimentale (e.g., Cohen, 1981 ; Gigerenzer, 1991, 2008) et ceux qui la défendent (e.g., Shafir & LeBoeuf, 2002 ; Stanovich et al., 2016).

Un argument souvent avancé par les opposants à la rationalité normative est que d'autres critères de rationalité permettraient d'évaluer le raisonnement des individus (Wagner-Egger, 2011). A titre d'exemple, on pourrait accorder davantage de pertinence à la rationalité écologique (Evans, 2007), qui évalue dans quelle mesure le comportement d'une personne est adapté à son environnement lorsqu'il s'agit d'évaluer le comportement humain. Selon cette définition de la rationalité, nous aurions pu considérer comme rationnel, dans la plupart des cas, que les individus désignent les individus blâmables au reste de la société pour maintenir une certaine harmonie sociale. Néanmoins, cet argument ne tient pas lorsque les déterminants de cet effet sont notamment liés au groupe d'appartenance de l'individu (e.g., origine ethnique, genre). Nous aurions également pu utiliser le raisonnement bayésien comme modèle normatif (Hagmayer, 2016). Le raisonnement bayésien consiste à ajuster les probabilités a priori (e.g., probabilité qu'un agent cause un dommage) avec les nouvelles informations obtenues sur le monde. Dans cette optique, les individus peuvent considérer que la probabilité qu'un agent

DISCUSSION GÉNÉRALE

blâmable provoque un dommage est plus élevée que celle d'un agent louable. Cette probabilité a priori plus élevée pour l'agent blâmable serait liée à la connaissance selon laquelle, en moyenne, les individus malveillants provoquent davantage de dommages que les individus bienveillants. Par conséquent, en recevant les mêmes informations sur l'événement, les individus attribueraient un score de causalité plus élevé à l'agent blâmable en raison d'une probabilité a priori plus élevée. Pour les mêmes raisons, les participants ont pu considérer que le fait que l'agent blâmable (vs. louable) ait causé le dommage est une explication plus plausible à fournir (Le Floch, 2008). Enfin, la rationalité normative que nous avons mobilisée est liée à la rationalité épistémique, à savoir le degré de correspondance entre nos croyances et la structure réelle du monde (Stanovich et al., 2016). Or, il paraît quelque peu idéaliste de considérer que les théories philosophiques actuelles modélisent parfaitement la « véritable » causalité, si une telle chose existe.

Ce débat et ces critiques sur la rationalité normative sont légitimes mais, comme nous l'avons abordé en Introduction Générale, la psychologie expérimentale peut également avoir des objectifs pratiques et des applications sociétales. Or, les deux critères normatifs que nous avons mobilisés présentent une efficacité opérationnelle certaine. Lorsque les ingénieurs de la NASA ont appris que la faible étanchéité des joints toriques d'un des propulseurs de la navette spatiale Challenger était une condition nécessaire réelle de son explosion, ils ont pu agir sur la confection de ces joints pour éviter d'autres accidents (Hilton et al., 1992). Le critère de nécessité réelle, comme celui de sensibilité, permet aux individus d'être efficaces face à leur environnement. Cette efficacité peut laisser suggérer que ces critères correspondent alors à quelque chose d'objectif dans notre environnement. Par conséquent, si des aspects de ces critères peuvent-être philosophiquement débattus, leur utilisation comme critères normatifs pour évaluer les jugements humains n'est pas totalement dénué d'intérêt. Par ailleurs, comme le soulignent Stanovich et ses collaborateurs (2016), les individus adhèrent rétrospectivement

DISCUSSION GÉNÉRALE

aux critères normatifs qu'ils ont violés. Or, le fait que les individus approuvent ces critères lorsqu'on les leur présente explicitement suggèrent qu'ils reconnaissent la force normative de ces règles de la rationalité (p.281). De plus, les théories juridiques de l'équivalence des conditions et de la causalité adéquate semblent suffisamment efficaces pour garantir une certaine justice et ne pas condamner des individus qui n'ont pas contribué à la survenue d'un dommage.

En somme, nos travaux ne remplissent qu'une partie des critères de Wagner-Egger (2011) pour affirmer que l'effet de la valence morale est dû à un biais cognitif. Néanmoins, nous avons apporté un certain nombre d'éléments novateurs permettant de faire avancer le débat : des critères normatifs pertinents et efficaces, bien que discutables, sont connus et mobilisés par les participants ; l'effet de la valence morale semble dépasser le problème d'interprétation de la tâche de jugement causal.

De la mesure des évaluations spontanées à la généralisation des résultats

Enfin, nous allons également discuter des limites concernant la validité des construits (i.e., capacité à opérationnaliser les concepts théoriques étudiés) et la validité externe (i.e., capacité à généraliser nos résultats à l'ensemble de la population humaine) de nos travaux ainsi que de l'ensemble des études sur l'effet de la valence morale.

Validité des construits. Nous avons observé une première limite qui concerne la mesure des évaluations spontanées négatives envers l'agent. Ces évaluations sont censées, selon le CCM, initier et modérer le *processus de validation du blâme*. Comme toutes les études sur ce processus (Alicke, 1992 ; Alicke et al., 2008, 2011 ; Lagnado & Channon, 2008 ; Rogers et al., 2019), nous avons utilisé des mesures subjectives auto-rapportées (e.g., Comment évaluez-vous la décision de Barbara de vouloir tuer Jean ?). L'opérationnalisation des motivations des agents de nos récits semblait avoir fonctionné puisqu'il y avait de très larges

DISCUSSION GÉNÉRALE

tailles d'effet des motivations sur les scores d'évaluation. Néanmoins, l'évaluation morale peut être provoquée par un jugement dénué d'émotion de colère (Pettigrove & Tanaka, 2014). Pour déterminer si ces évaluations sont, au moins en partie, liées à des réactions émotionnelles conformément au CCM, il pourrait être pertinent d'utiliser des mesures physiologiques telles que l'imagerie cérébrale (Greene et al., 2001) ou la réaction électrodermale (Sequeira & D'Hondt, 2013). Ces mesures pourraient également permettre de départager le CCM des modèles du raisonnement contrefactuel, ces derniers considérant que le raisonnement contrefactuel se focalise sur la condition anormale sans aucune implication émotionnelle.

De plus, nous avons étudié l'effet de la valence morale avec des facteurs légalement pertinents pour l'attribution d'une responsabilité et de sanctions, tels que des états mentaux légalement répréhensibles. Or, les évaluations spontanées négatives peuvent survenir en raison de facteurs extralégaux comme la réputation de l'agent, son attractivité sociale, son origine ethnique ou son genre (Alicke, 2000 ; Alicke et al., 1990, 2008 ; Alicke & Zell, 2009). Il pourrait être intéressant d'étudier l'effet de la valence morale avec des facteurs extralégaux de ce type. Plus précisément, nous pourrions examiner l'effet de la valence morale en manipulant les deux dimensions fondamentales de la perception sociale, à savoir la chaleur et la compétence (Fiske et al., 2007). Par exemple, il serait possible de manipuler le groupe d'appartenance de l'agent pour susciter des émotions négatives ou positives. Selon le Modèle du Contenu des Stéréotypes (Fiske et al., 2002), les individus appartenant à des groupes associés à une faible chaleur et une forte compétence (e.g., les chefs d'entreprise, les personnes riches) susciteraient l'émotion négative de l'envie ; tandis que les individus appartenant à des groupes associés à une forte chaleur et une faible compétence (e.g., personnes présentant un handicap, personnes âgées) pourraient être associés à l'émotion positive de la pitié. Si un effet du groupe d'appartenance sur le jugement causal est observé, cela renforcerait encore davantage le CCM au détriment de l'approche pragmatique. En effet, un tel résultat indiquerait

DISCUSSION GÉNÉRALE

que cet effet serait lié à un raisonnement motivé par une évaluation négative, plutôt qu'un raisonnement rationnel et une interprétation de la question en termes de responsabilité.

Une autre limite liée aux concepts théoriques est que les travaux présentant un agent louable et un agent blâmable ne permettent pas de déterminer avec certitude si l'effet de la valence morale observé est intégralement lié à un *processus de validation du blâme* (Alicke, 2000), ou si une partie de l'effet s'explique par une *validation de l'excuse* (Turri & Blouw, 2014), à savoir minimiser le rôle causal de l'agent louable pour éviter de le blâmer. Par conséquent, il pourrait être pertinent de mener une étude avec un groupe contrôle (e.g., un agent qui provoque le dommage de façon accidentelle et non négligente) pour étudier si c'est le score de l'agent blâmable qui augmente, celui de l'agent louable qui diminue, ou les deux à la fois.

Validité externe. Tout comme plus de 80% des travaux en psychologie expérimentale (Henrich et al., 2010), nous ignorons dans quelle mesure nos résultats peuvent se généraliser au-delà des sociétés WEIRD (Western, Educated, Industrialized, Rich and Democratic). Néanmoins, concernant l'effet de la valence morale, des variations interculturelles dans les normes morales des sociétés ne devraient pas avoir d'impact sur le *processus de validation du blâme*. En effet, il suffit qu'un agent soit évalué négativement, indépendamment des valeurs morales de l'individu qui juge, pour que la surestimation de son rôle causal ait lieu. Il suffit alors d'adapter la valence morale de l'agent en fonction des normes morales de la culture d'appartenance des participants. En effet, ce qui importe dans la validation du blâme est l'évaluation négative envers l'agent, et ce, peu importe la source de cette évaluation. Par ailleurs, il convient de noter que l'effet de la valence morale a été répliqué au sein d'une société non-WEIRD, avec des participants indiens (Kanekar et al., 1993). Néanmoins, la question des variations interculturelles pourrait davantage se poser pour le respect des critères normatifs.

DISCUSSION GÉNÉRALE

En effet, les théories du raisonnement contrefactuel supposent que l'individu se base sur ses connaissances antérieures pour effectuer des simulations contrefactuelles. Ces simulations permettent alors de déterminer si un agent est une condition nécessaire réelle ou non, ainsi que son degré de sensibilité aux circonstances. Dans le cas où les individus d'une culture donnée croient en la « pensée magique » (i.e., provoquer des événements par la seule force de la pensée, voir Gardair, 2014), que se passerait-il vis-à-vis du lien causal préempté (i.e., l'agent souhaitait provoquer le dommage) ? Les individus iraient-ils jusqu'à considérer que les intentions de Barbara sont une condition nécessaire réelle au décès de son mari ? LeGuen et ses collègues (2015) ont effectivement observé que les participants issus de cultures croyant en la « pensée magique » (mayas Yucatec et Tzeltal) attribuaient un score de causalité plus élevé à l'agent en l'absence d'un lien causal, par rapport aux participants issus de cultures WEIRD (allemands). Néanmoins, ces mêmes participants mayas Yucatec et Tzeltal attribuaient à l'agent un score de causalité considérablement réduit, comparativement à la condition présence d'un lien causal, témoignant d'une intégration et d'une application du critère de nécessité réelle. L'étude présentait toutefois des situations à faible gravité (e.g., fenêtre cassée, mort d'un insecte). De futures recherches pourraient étudier le respect de ces critères dans un contexte favorable au *processus de validation du blâme* pour des individus croyant en la « pensée magique ».

Conclusion

Après une trentaine d'années de recherche et de répliques, l'effet de la valence morale sur le jugement causal semble porter atteinte à la rationalité du sens commun. Or, la rationalité des jugements causaux est cruciale pour s'adapter efficacement à l'environnement et pour éviter des conséquences sociétales indésirables, notamment à travers des décisions judiciaires partiales. Malgré leurs limites, nos travaux ont permis d'éclaircir le débat en situant l'effet de la valence morale par rapport à des critères normatifs philosophiques (i.e., nécessité réelle et sensibilité) et pragmatiques (i.e., maximes de relation et de quantité). Notre premier axe de recherche a permis de démontrer que les individus connaissent et appliquent les deux critères normatifs de la causalité, et ce, de façon prioritaire. Par conséquent, un agent ne sera pas désigné comme cause d'un dommage si ses actions ne sont pas des conditions nécessaires réelles à celui-ci. De plus, le critère de sensibilité semble également occuper une place centrale dans les jugements causaux de sens commun, celui-ci étant appliqué prioritairement par une majorité d'individus. Ces deux affirmations sont valables même lorsque les individus semblent fortement motivés à blâmer et à sanctionner un agent. Notre second axe de recherche, quant à lui, a corroboré la place centrale du critère de sensibilité, mais a réfuté l'hypothèse pragmatique selon laquelle l'effet de la valence morale serait lié à une interprétation de la tâche comme une recherche de responsable (i.e., *enquête attributive*). Par conséquent, bien que les critères normatifs de la causalité soient centraux, l'apparition de l'effet de la valence morale serait probablement due à un biais cognitif, conformément au CCM et contrairement à l'approche pragmatique du jugement causal.

Les jugements causaux de sens commun semblent donc effectivement biaisés par une motivation à blâmer, mais uniquement dans des circonstances délimitées par les deux critères normatifs de la causalité et lorsque l'évaluation envers l'agent est suffisamment négative. Ces découvertes ouvrent la voie à de futures investigations visant à approfondir davantage notre

DISCUSSION GÉNÉRALE

compréhension de l'effet de la valence morale, tant du point de vue des mécanismes sous-jacents que des facteurs sociaux qui y sont associés. Par ailleurs, nous espérons que ces travaux constitueront une base pour guider de futures études sur le débiaisement du *processus de validation du blâme*. Nos résultats suggèrent notamment de tester des méthodes en lien avec l'inhibition ou la détection du conflit pour les quelques cas où l'effet de la valence morale apparaît, tout en prenant en compte de potentielles différences inter-individuelles telles que le locus de contrôle, l'identification à l'agent ou la croyance en la pensée magique.

Références

- Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 63(3), 368–378. <https://doi.org/10.1037/0022-3514.63.3.368>
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126(4), 556–574. <https://doi.org/10.1037/0033-2909.126.4.556>
- Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable Control and Counterfactual Reasoning in the Psychology of Blame. *Personality and Social Psychology Bulletin*, 34(10), 1371–1381. <https://doi.org/10.1177/0146167208321594>
- Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., & Lagnado, D. A. (2015). Causal Conceptions in Social Explanation and Moral Evaluation: A Historical Tour. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 10(6), 790–812. <https://doi.org/10.1177/1745691615601888>
- Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, Norm violation, and Culpable control. *The Journal of Philosophy*, 108(12), 670-696. JSTOR.
- Alicke, M. D., & Zell, E. (2009). Social Attractiveness and Blame. *Journal of Applied Social Psychology*, 39(9), 2089-2105. <https://doi.org/10.1111/j.1559-1816.2009.00517.x>
- Andreas, H., Armgardt, M., & Gunther, M. (2022). Counterfactuals for causal responsibility in legal contexts. *Artificial Intelligence and Law*, 31, 115 - 132. <https://doi.org/10.1007/s10506-021-09307-2>
- Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The Impact of Jury Race in Criminal Trials. *The Quarterly Journal of Economics*, 127(2), 1017-1055. <https://doi.org/10.1093/qje/qjs014>

- Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*, 486-498. <https://doi.org/10.1037/0033-2909.110.3.486>
- Arkkelin, D., Oakley, T., & Mynatt, C. R. (1979). Effects of controllable versus uncontrollable factors on responsibility attributions: A single-subject approach. *Journal of Personality and Social Psychology*, *37*(1), 110-115. <https://doi.org/10.1037/0022-3514.37.1.110>
- Arnaud, P., & Ouss, A. (2016). L'impact des médias sur les décisions de justice. *Notes IPP*, *22*. <https://shs.hal.science/halshs-02522870>
- Austin, J. L. (1962). *How to Do Things with Words* (M. Sbisá & J. O. Urmson, Éds.). Oxford: Clarendon Press.
- Babai, R., Shalev, E., & Stavy, R. (2015). A warning intervention improves students' ability to overcome intuitive interference. *ZDM*, *47*(5), 735–745. <https://doi.org/10.1007/s11858-015-0670-y>
- Bachrach, R., Huesmann, L. R., & Peterson, R. A. (1977). The Relation between Locus of Control and the Development of Moral Judgment. *Child Development*, *48*(4), 1340-1352. <https://doi.org/10.2307/1128492>
- Bago, B., & De Neys, W. (2017). Fast logic?: Examining the time course assumption of dual process theory. *Cognition*, *158*, 90-109. <https://doi.org/10.1016/j.cognition.2016.10.014>
- Bago, B., & De Neys, W. (2019). The Smart System 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *25*(3), 257–299. <https://doi.org/10.1080/13546783.2018.1507949>
- Bang, D., & Frith, C. D. (2017). Making better decisions in groups. *Royal Society Open Science*, *4*(8), 170193. <https://doi.org/10.1098/rsos.170193>

- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. <https://doi.org/10.1037/0033-2909.128.4.612>
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59–78. <https://doi.org/10.1017/S0140525X11002202>
- Beebe, H., Hitchcock, C., & Menzies, P. (2009). *The Oxford Handbook of Causation*. Oxford University Press UK.
- Berger, E. (2008). *La justice pénale sous la Révolution : Les enjeux d'un modèle judiciaire libéral*. Presses universitaires de Rennes.
- Bernicot, J. (2002). *Pragmatique et psychologie*. Presses universitaires de Nancy.
- Berthet, V. (2021). The Measurement of Individual Differences in Cognitive Biases: A Review and Improvement. *Frontiers in Psychology*, *12*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.630177>
- Blanchette, I., Caparos, S., & Trémolière, B. (2018). Emotion and reasoning. In L. J. Ball & V. A. Thompson (Eds.), *The Routledge international handbook of thinking and reasoning* (pp. 57–70). Routledge/Taylor & Francis Group.
- Bloom, P. (2006). My Brain Made Me Do It. *Journal of Cognition and Culture*, *6*(1-2): 1567–7095. <https://doi.org/10.1163/156853706776931303>

- Bloom, P. (2013). *Just babies: The origins of good and evil*. Crown Publishers/Random House.
- Bordel, S. (2002). *Les dimensions objective et subjective du jugement de responsabilité* [Thèse de doctorat, Université Rennes 2]. <https://www.theses.fr/2002REN20038>
- Bordel, S., Guingouain, G., & Somat, A. (2006). Objective and Subjective Responsibility in a Judicial Context. *Swiss Journal of Psychology*, 65(4), 227-235. <https://doi.org/10.1024/1421-0185.65.4.227>
- Bornstein, B. H., Golding, J. M., Neuschatz, J., Kimbrough, C., Reed, K., Magyarics, C., & Luecht, K. (2017). Mock juror sampling issues in jury simulation research: A meta-analysis. *Law and Human Behavior*, 41(1), 13–28. <https://doi.org/10.1037/lhb0000223>
- Bowers, R. I. (2021). Causal Reasoning. In T. K. Shackelford & V. A. Weekes-Shackelford (Éds.), *Encyclopedia of Evolutionary Psychological Science* (pp. 920-936). Springer International Publishing. https://doi.org/10.1007/978-3-319-19650-3_3114
- Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. In H. W. Reese & R. Kail (Eds.), *Advances in child development and behavior* (pp. 41-100). Academic Press.
- Brickman, P., Ryan, K., & Wortman, C. B. (1975). Causal chains: Attribution of responsibility as a function of immediate and prior causes. *Journal of Personality and Social Psychology*, 32, 1060-1067. <https://doi.org/10.1037/0022-3514.32.6.1060>
- Buehner, M. J. (2017). Space, Time, and Causality. In M. R. Waldmann (Éd.), *The Oxford Handbook of Causal Reasoning* (pp. 549-564). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199399550.013.29>

- Burger, J. M. (1981). Motivational biases in the attribution of responsibility for an accident: A meta-analysis of the defensive-attribution hypothesis. *Psychological Bulletin*, *90*, 496-512. <https://doi.org/10.1037/0033-2909.90.3.496>
- Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, *36*, 1311-1321. <https://doi.org/10.1037/0022-3514.36.11.1311>
- Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, *6*(10), 426-431. [https://doi.org/10.1016/S1364-6613\(02\)01974-5](https://doi.org/10.1016/S1364-6613(02)01974-5)
- Byrne, R. M. J. (2007). Précis of The Rational Imagination: How People Create Alternatives to Reality. *Behavioral and Brain Sciences*, *30*(5-6), 439-452. <https://doi.org/10.1017/S0140525X07002579>
- Byrne, R. M. J. (2016). Counterfactual Thought. *Annual Review of Psychology*, *67*(1), 135-157. <https://doi.org/10.1146/annurev-psych-122414-033249>
- Byrne, R. M. J., & Johnson-Laird, P. N. (2020). If and or: Real and counterfactual possibilities in their truth and probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 760-780. <https://doi.org/10.1037/xlm0000756>
- Cane, P. (2002). *Responsibility in Law and Morality*. Bloomsbury Publishing.
- Carlson, R. W., Bigman, Y. E., Gray, K., Ferguson, M. J., & Crockett, M. J. (2022). How inferred motives shape moral judgements. *Nature Reviews Psychology*, *1*(8), 468–478. <https://doi.org/10.1038/s44159-022-00071-x>

- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, *12*(3), 314-325. <https://doi.org/10.1006/nimg.2000.0612>
- Caverni, J.-P., Fabre, J.-M., & Gonzalez, M. (1990). *Cognitive Biases*. Elsevier.
- Charon, J.-M. (2008). Le traitement médiatique de l'affaire d'Outreau. *Droit et cultures. Revue internationale interdisciplinaire*, *55*, 221-239. <https://doi.org/10.4000/droitcultures.1387>
- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287-291. <https://doi.org/10.1016/j.tics.2006.05.007>
- Chen, H., Cohen, P., & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*, *39*(4), 860-864. <https://doi.org/10.1080/03610911003650383>
- Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology*, *58*(4), 545-567. <https://doi.org/10.1037/0022-3514.58.4.545>
- Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*(2), 365-382. <https://doi.org/10.1037/0033-295X.99.2.365>
- Clarke, R., Shepherd, J., Stigall, J., Waller, R. R., & Zarpentine, C. (2015). Causation, norms, and omissions: A study of causal judgments. *Philosophical Psychology*, *28*(2), 279–293. <https://doi.org/10.1080/09515089.2013.815099>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4(3), 317-331. <https://doi.org/10.1017/S0140525X00009092>
- Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and Causation: History, Problems, and Prospects. In J. Collins, N. Hall, & L. Paul (Éds.), *Causation and Counterfactuals* (pp. 1-57). MIT Press.
- Cooper, G. (1999). An Overview of the Representation and Discovery of Causal Relationships Using Bayesian Networks. Dans *Computation, causation, and discovery* (pp. 4-62). <https://doi.org/10.7551/mitpress/2006.003.0002>
- Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is It Good to Cooperate?: Testing the Theory of Morality-as-Cooperation in 60 Societies. *Current Anthropology*, 60(1), 47-69. <https://doi.org/10.1086/701478>
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108(2), 353-380. <https://doi.org/10.1016/j.cognition.2008.03.006>
- De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(1), 28-38. <https://doi.org/10.1177/1745691611429354>
- De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking and Reasoning*, 20(2), 169-187. <https://doi.org/10.1080/13546783.2013.854725>

- De Neys, W., & Bonnefon, J.-F. (2013). The ‘whys’ and ‘whens’ of individual differences in thinking biases. *Trends in Cognitive Sciences*, 17(4), 172-178.
<https://doi.org/10.1016/j.tics.2013.02.001>
- De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, 113(1), 45-61.
<https://doi.org/10.1016/j.cognition.2009.07.009>
- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28(5), 503-509.
<https://doi.org/10.1177/0963721419855658>
- Debove, S., André, J.-B., & Baumard, N. (2015). Partner choice creates fairness in humans. *Proceedings of the Royal Society B: Biological Sciences*, 282(1808), 20150392.
<https://doi.org/10.1098/rspb.2015.0392>
- Decety, J., & Cacioppo, S. (2012). The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology*, 108(11), 3068-3072.
<https://doi.org/10.1152/jn.00473.2012>
- Demirtas, H. (2022). Causation comes in degrees. *Synthese*, 200(2), 64.
<https://doi.org/10.1007/s11229-022-03507-2>
- Devine, D. J., Clayton, L. D., Dunford, B. B., Seying, R., & Pryce, J. (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, Public Policy, and Law*, 7(3), 622. <https://doi.org/10.1037/1076-8971.7.3.622>
- Devos-Comby, L., & Devos, T. (2001). Social norms, social value, and judgments of responsibility. *Swiss Journal of Psychology / Schweizerische Zeitschrift für Psychologie*

/ *Revue Suisse de Psychologie*, 60(1), 35-46. <https://doi.org/10.1024/1421-0185.60.1.35>

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5, 781. <https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00781>

Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous Skepticism: The Interplay of Motivation and Expectation in Responses to Favorable and Unfavorable Medical Diagnoses. *Personality and Social Psychology Bulletin*, 29(9), 1120-1132. <https://doi.org/10.1177/0146167203254536>

Dowe, P. (2000). *Physical Causation*. Cambridge University Press.

Driver, J. (2008). Attributions of causation and moral responsibility. In *Moral psychology, Vol 2: The cognitive science of morality: Intuition and diversity* (pp. 423-439). MIT Press.

Dror, I. E. (2018). Biases in forensic experts. *Science*, 360(6386), 243-243. <https://doi.org/10.1126/science.aat8443>

Drouet, I. (2007). *Causalité et probabilités : Réseaux bayésiens, propensionnisme* [Thèse de doctorat, Université Panthéon-Sorbonne - Paris I]. <https://www.theses.fr/2007PA010679>

Dyczewski, E. A., & Markman, K. D. (2012). General attainability beliefs moderate the motivational effects of counterfactual thinking. *Journal of Experimental Social Psychology*, 48(5), 1217-1220. <https://doi.org/10.1016/j.jesp.2012.04.016>

Efran, M. G. (1974). The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality*, 8(1), 45-54. [https://doi.org/10.1016/0092-6566\(74\)90044-0](https://doi.org/10.1016/0092-6566(74)90044-0)

- Ellis, G. (2018). So, What Are Cognitive Biases? In G. Ellis (Éd.), *Cognitive Biases in Visualizations* (pp. 1-10). Springer International Publishing.
https://doi.org/10.1007/978-3-319-95831-6_1
- Engelmann, N., & Waldmann, M. R. (2021). *A Causal Proximity Effect in Moral Judgment* [Conférence]. Proceedings of the Annual Meeting of the Cognitive Science Society, Vienne (Autriche). <https://escholarship.org/uc/item/9hp8q72s>
- Engelmann, N., & Waldmann, M. R. (2022). How causal structure, causal strength, and foreseeability affect moral judgments. *Cognition*, 226, 105167.
<https://doi.org/10.1016/j.cognition.2022.105167>
- Etzioni, A. (2014). Treating Rationality as a Continuous Variable. *Society*, 51(4), 393-400.
<https://doi.org/10.1007/s12115-014-9798-6>
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255-278.
<https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. S. B. T. (2019). *Hypothetical Thinking: Dual Processes in Reasoning and Judgement*. Psychology Press.
- Evans, J. S. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology*, 6(3), 263-285.
<https://doi.org/10.1080/09541449408520148>
- Evans, J. St. B. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Psychology Press.

- Feigenson, N. (2016). Jurors' Emotions and Judgments of Legal Responsibility and Blame: What Does the Experimental Research Tell Us? *Emotion Review*, 8(1), 26-31. <https://doi.org/10.1177/1754073915601223>
- Feigenson, N., & Park, J. (2006). Emotions and Attributions of Legal Responsibility and Blame: A Research Review. *Law and Human Behavior*, 30(2), 143-161. <https://doi.org/10.1007/s10979-006-9026-z>
- Fenker, D. B., Waldmann, M. R., & Holyoak, K. J. (2005). Accessing causal relations in semantic memory. *Memory & Cognition*, 33(6), 1036-1046. <https://doi.org/10.3758/BF03193211>
- Fincham, F. D., & Jaspars, J. M. (1980). Attribution of Responsibility: From Man the Scientist to Man As Lawyer. In L. Berkowitz (Éd.), *Advances in Experimental Social Psychology*, 13, 81-138. Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60131-8](https://doi.org/10.1016/S0065-2601(08)60131-8)
- Fincham, F. D., & Jaspars, J. M. (1983). A subjective probability approach to responsibility attribution. *British Journal of Social Psychology*, 22(2), 145-161. <https://doi.org/10.1111/j.2044-8309.1983.tb00575.x>
- Fincham, F. D., & Roberts, C. (1985). Intervening causation and the mitigation of responsibility for harm doing II. The role of limited mental capacities. *Journal of Experimental Social Psychology*, 21(2), 178-194. [https://doi.org/10.1016/0022-1031\(85\)90014-9](https://doi.org/10.1016/0022-1031(85)90014-9)
- Fincham, F. D., & Shultz, T. R. (1981). Intervening causation and the mitigation of responsibility for harm. *British Journal of Social Psychology*, 20(2), 113-120. <https://doi.org/10.1111/j.2044-8309.1981.tb00483.x>

- Finkel, N. J., & Groscup, J. L. (1997). Crime Prototypes, Objective Versus Subjective Culpability, and a Commonsense Balance. *Law and Human Behavior*, 21(2), 209-230. <https://doi.org/10.1023/A:1024830413404>
- Fischer, G.-N. (2020). Chapitre 5. La cognition sociale. In *Les concepts fondamentaux de la psychologie sociale* (pp. 159-192). Dunod. <https://doi.org/10.3917/dunod.fisch.2020.01.0159>
- Fischhoff, B. (1982). *Debiasing*. In Daniel Kahneman, Paul Slovic & Amos Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809477.032>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77-83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Fletcher, L., & Carruthers, P. (2012). Metacognition and reasoning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1366-1378. <https://doi.org/10.1098/rstb.2011.0413>
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25-42. <https://doi.org/10.1257/089533005775196732>

- Frey, D., Johnson, E. D., & Neys, W. D. (2018). Individual differences in conflict detection during reasoning. *Quarterly Journal of Experimental Psychology*, *71*(5), 1188-1208. <https://journals.sagepub.com/doi/10.1080/17470218.2017.1313283>
- Fritz, M. S., & MacKinnon, D. P. (2007). Required Sample Size to Detect the Mediated Effect. *Psychological science*, *18*(3), 233-239. <https://doi.org/10.1111/j.1467-9280.2007.01882.x>
- Fumerton, R., & Kress, K. (2001). Causation and the Law: Preemption, Lawful Sufficiency, and Causal Sufficiency. *Law and Contemporary Problems*, *64*(4), 83-105. <https://doi.org/10.2307/1192292>
- Gailey, J. A., & Falk, R. F. (2008). Attribution of Responsibility as a Multidimensional Concept. *Sociological Spectrum*, *28*(6), 659-680. <https://doi.org/10.1080/02732170802342958>
- Gardair, E. (2014). Cognitions « rationnelles » et « irrationnelles » : quels liens entretiennent la pensée magique et la pensée scientifique dans les raisonnements quotidiens ? *Intellectica*, *62*, 51-68. https://www.persee.fr/doc/intel_0769-4113_2014_num_62_2_1033
- Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*(5), 1012-1020. <https://doi.org/10.1111/j.1467-7687.2011.01048.x>
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165-193. [https://doi.org/10.1016/0010-0277\(95\)00661-H](https://doi.org/10.1016/0010-0277(95)00661-H)

- Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(5), 936-975. <https://doi.org/10.1037/rev0000281>
- Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-Tracking Causality. *Psychological Science*, *28*(12), 1731-1744. <https://doi.org/10.1177/0956797617713053>
- Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond “Heuristics and Biases”. *European Review of Social Psychology*, *2*(1), 83-115. <https://doi.org/10.1080/14792779143000033>
- Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science*, *3*(1), 20-29. <https://doi.org/10.1111/j.1745-6916.2008.00058.x>
- Gilbert, E. A., Tenney, E. R., Holland, C. R., & Spellman, B. A. (2015). Counterfactuals, control, and causation: Why knowledgeable people get blamed more. *Personality & Social Psychology Bulletin*, *41*(5), 643-658. <https://doi.org/10.1177/0146167215572137>
- Gissinger-Bosse, C. (2012). *Vers une conversion démocratique : Analyse du dispositif de parole de la cour d'assises* [Thèse de doctorat, Strasbourg]. <https://www.theses.fr/2012STRAG018>
- Glymour, C., & Cheng, P. W. (1998). Causal Mechanism and Probability: A Normative Approach. Dans Mike Oaksford and Nick Chater (Eds), *Rational Models of Cognition*. Oxford University Press. <https://doi.org/10.1184/R1/6491081.v1>

- Goldinger, S. D., Kleider, H. M., Azuma, T., & Beike, D. R. (2003). "Blaming The Victim" Under Memory Load. *Psychological Science*, 14(1), 81-85.
<https://doi.org/10.1111/1467-9280.01423>
- Goldvarg, E., & Johnson-Laird, P. n. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, 25(4), 565-610.
https://doi.org/10.1207/s15516709cog2504_3
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371. <https://doi.org/10.1037/0033-295x.101.3.371>
- Green, S. (2015). *Causation in Negligence*. Bloomsbury Publishing.
- Greenawalt, K. (1986). Distinguishing Justifications from Excuses. *Law and Contemporary Problems*, 49, 89.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293(5537), 2105-2108. <https://doi.org/10.1126/science.1062872>
- Grice, H. P. (1975). Logic and Conversation. Dans P.Cole & J.L. Morgan (Eds), *Syntax and semantics 3: Speech Acts*. New York: Wiley.
https://doi.org/10.1163/9789004368811_003
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, 51(4), 334-384. <https://doi.org/10.1016/j.cogpsych.2005.05.004>
- Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging Levels of Analysis for Probabilistic Models of Cognition. *Current Directions in Psychological Science*, 21(4), 263-268.
<https://doi.org/10.1177/0963721412447619>

- Grinfeld, G., Lagnado, D., Gerstenberg, T., Woodward, J. F., & Usher, M. (2020). Causal Responsibility and Robust Causation. *Frontiers in Psychology*, *11*, 1069. <https://doi.org/10.3389/fpsyg.2020.01069>
- Gross, J. J., Richards, J. M., & John, O. P. (2006). Emotion Regulation in Everyday Life. Dans *Emotion regulation in couples and families: Pathways to dysfunction and health* (pp. 13-35). American Psychological Association. <https://doi.org/10.1037/11468-001>
- Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the Heart of Morality Lies Folk Psychology. *Inquiry*, *52*(5), 449-466. <https://doi.org/10.1080/00201740903302600>
- Güver, L., & Kneer, M. (2022). Causation and the Silly Norm Effect. *Advances in Experimental Philosophy of Law*. (Forthcoming). <https://papers.ssrn.com/abstract=4047203>
- Hagmayer, Y. (2016). Causal Bayes nets as psychological theories of causal reasoning: Evidence from psychological research. *Synthese*, *193*(4), 1107-1126. <https://doi.org/10.1007/s11229-015-0734-0>
- Hall, N. (2004). Two Concepts of Causation. In J. Collins, N. Hall, & L. Paul (Eds.), *Causation and Counterfactuals* (pp. 225-276). MIT Press.
- Halpern, J. Y. (2016). *Actual Causality*. The MIT Press; JSTOR. <https://doi.org/10.2307/j.ctt1f5g5p9>
- Halpern, J. Y., & Hitchcock, C. (2015). Graded Causation and Defaults. *The British Journal for the Philosophy of Science*, *66*(2), 413-457. <https://doi.org/10.1093/bjps/axt050>
- Halpern, J. Y., & Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *The British Journal for the Philosophy of Science*, *56*(4), 843-887. <https://doi.org/10.1093/bjps/axi147>

- Hamilton, V. L. (1978). Who is Responsible? Toward a Social Psychology of Responsibility Attribution. *Social Psychology*, 41(4), 316-328. <https://doi.org/10.2307/3033584>
- Hamilton, V. L. (1980). Intuitive psychologist or intuitive lawyer? Alternative models of the attribution process. *Journal of Personality and Social Psychology*, 39(5), 767-772. <https://doi.org/10.1037/0022-3514.39.5.767>
- Hamlin, J. K. (2013). Moral Judgment and Action in Preverbal Infants and Toddlers: Evidence for an Innate Moral Core. *Current Directions in Psychological Science*, 22(3), 186-193. <https://doi.org/10.1177/0963721412470687>
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450(7169), 557-559. <https://doi.org/10.1038/nature06288>
- Hart, H. L. A., & Honoré, T. (1985). *Causation in the Law* (2nd ed.). OUP Oxford.
- Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, 41(2), 129-140. <https://doi.org/10.1016/j.jesp.2004.06.008>
- Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology*, 46(1), 44-56. <https://doi.org/10.1037/0022-3514.46.1.44>
- Hastie, R. (1993). *Inside the Juror: The Psychology of Juror Decision Making*. Cambridge University Press.
- Hayes, A. F., & Scharkow, M. (2013). The Relative Trustworthiness of Inferential Tests of the Indirect Effect in Statistical Mediation Analysis: Does Method Really Matter? *Psychological Science*, 24(10), 1918-1927. <https://doi.org/10.1177/0956797613480187>
- Heider, F. (1946). Attitudes and Cognitive Organization. *The Journal of Psychology*, 21(1), 107-112. <https://doi.org/10.1080/00223980.1946.9917275>

- Heider, F. (1958). *The psychology of interpersonal relations*. John Wiley & Sons Inc.
<https://doi.org/10.1037/10628-000>
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243-259. <https://doi.org/10.2307/1416950>
- Henne, P., Kulesza, A., Perez, K., & Houcek, A. (2021). Counterfactual thinking and recency effects in causal judgment. *Cognition*, 212, 104708.
<https://doi.org/10.1016/j.cognition.2021.104708>
- Henne, P., Pinillos, Á., & De Brigard, F. (2017). Cause by Omission and Norm: Not Watering Plants. *Australasian Journal of Philosophy*, 95(2), 270-283.
<https://doi.org/10.1080/00048402.2016.1182567>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83.
<https://doi.org/10.1017/S0140525X0999152X>
- Herbert, T. T., & Estes, R. W. (1977). Improving Executive Decisions by Formalizing Dissent: The Corporate Devil's Advocate. *The Academy of Management Review*, 2(4), 662-667.
<https://doi.org/10.2307/257518>
- Hesslow, G. (1988). The problem of causal selection. Dans D.J. Hilton (Ed.), *Contemporary science and natural explanation: Commonsense Conceptions of Causality* (pp. 11-32). New York University Press.
- Hilton, D. J. (2007). Causal explanation: From social perception to knowledge-based causal attribution. Dans A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 232-253). The Guilford Press.

- Hilton, D. J. (2017). Social attribution and explanation. Dans M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 645-674). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199399550.001.0001>
- Hilton, D. J. (1991). A Conversational Model of Causal Explanation. *European Review of Social Psychology*, 2(1), 51-81. <https://doi.org/10.1080/14792779143000024>
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, 118(2), 248-271. <https://doi.org/10.1037/0033-2909.118.2.248>
- Hilton, D. J. (1996). Mental Models and Causal Explanation: Judgements of Probable Cause and Explanatory Relevance. *Thinking & Reasoning*, 2(4), 273-308.
<https://doi.org/10.1080/135467896394447>
- Hilton, D. J. (2002). Le jugement de causalité et l'explication causale. Dans G. Politzer (Ed.), *Le raisonnement humain*. (pp. 217-239). Paris: Hermès.
- Hilton, D. J., Mathes, R.M. and Trabasso, T.R. (1992). The study of causal explanation in natural language: Analysing reports of the Challenger disaster in the "New York Times". Dans M. McLaughlin, S. Cody and S.J. Read (Eds.) *Explaining one's self to others: reason-giving in a social context*. Hillsdale, N.J.: Erlbaum.
- Hilton, D. J., McClure, J., & Moir, B. (2016). Acting knowingly: Effects of the agent's awareness of an opportunity on causal attributions. *Thinking & Reasoning*, 22(4), 461-494. <https://doi.org/10.1080/13546783.2016.1191547>
- Hilton, D. J., McClure, J., & Sutton, R. M. (2010). Selecting explanations from causal chains: Do statistical principles explain preferences for voluntary causes? *European Journal of Social Psychology*, 40(3), 383-400. <https://doi.org/10.1002/ejsp.623>

- Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, 93(1), 75-88. <https://doi.org/10.1037/0033-295X.93.1.75>
- Hilton, D. J., Schmeltzer, C., & Goulette, V. (2021). A question of detail: Matching counterfactuals to actual cause in pre-emption scenarios. *Thinking & Reasoning*, 27(3), 350-388. <https://doi.org/10.1080/13546783.2020.1810122>
- Hitchcock, C. (2007). Prevention, Preemption, and the Principle of Sufficient Reason. *The Philosophical Review*, 116(4), 495-532. <https://www.jstor.org/stable/20446988>
- Hitchcock, C., & Knobe, J. (2009). Cause and Norm. *The Journal of Philosophy*, 106(11), 587-612. JSTOR.
- Hoffman, R., Klein, G., & Miller, J. (2011). Naturalistic investigations and models of reasoning about complex indeterminate causation. *Information Knowledge Systems Management*, 10(1-4), 397-425. <https://doi.org/10.3233/IKS-2012-0203>
- Holvoet, C., Scola, C., Arciszewski, T., & Picard, D. (2016). Infants' preference for prosocial behaviors: A literature review. *Infant Behavior and Development*, 45, 125-139. <https://doi.org/10.1016/j.infbeh.2016.10.008>
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62, 135-163. <https://doi.org/10.1146/annurev.psych.121208.131634>
- Houdé, O. (2019). *3-System Theory of the Cognitive Brain: A Post-Piagetian Approach to Cognitive Development*. Routledge.
- Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., & Tzourio-Mazoyer, N. (2000). Shifting from the perceptual brain to the logical brain: The neural impact of

- cognitive inhibition training. *Journal of Cognitive Neuroscience*, 12(5), 721-728.
<https://doi.org/10.1162/089892900562525>
- Hume, D. (1748). *An enquiry concerning human understanding*. Oxford University Press..
- Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161, 80-93. <https://doi.org/10.1016/j.cognition.2017.01.010>
- Janoff-Bulman, R., Timko, C., & Carli, L. L. (1985). Cognitive biases in blaming the victim. *Journal of Experimental Social Psychology*, 21(2), 161-177.
[https://doi.org/10.1016/0022-1031\(85\)90013-7](https://doi.org/10.1016/0022-1031(85)90013-7)
- Jarosz, A., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, 7(1), 2-9.
<https://doi.org/10.7771/1932-6246.1167>
- Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. *Handbook of motivation and cognition: Foundations of social behavior, Vol. 2*. (pp. 482-526). The Guilford Press.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. Dans *Advances in Experimental Social Psychology*. Academic Press. 2, 219-266. [https://doi.org/10.1016/S0065-2601\(08\)60107-0](https://doi.org/10.1016/S0065-2601(08)60107-0)
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136-153. <https://doi.org/10.1037/0033-295X.93.2.136>
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-291. <https://doi.org/10.2307/1914185>

- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. Dans D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty*. Cambridge University Press. (pp. 201-208). <https://doi.org/10.1017/CBO9780511809477.015>
- Kanazawa, S. (1992). Outcome or Expectancy? Antecedent of Spontaneous Causal Attribution. *Personality and Social Psychology Bulletin*, 18(6), 659-668. <https://doi.org/10.1177/0146167292186001>
- Kanekar, S., Dhir, V. L., Franco, B., Sindhakar, A. R., Vaz, L., & Nazareth, A. M. (1993). Causality, Blame and Punishment. *The Irish Journal of Psychology*, 14(4), 596-604. <https://doi.org/10.1080/03033910.1993.10557974>
- Kapardis, A. (2009). *Psychology and Law: A Critical Introduction*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511813764>
- Karlovac, M., & Darley, J. M. (1988). Attribution of Responsibility for Accidents: A Negligence Law Analogy. *Social Cognition*, 6(4), 287-318. <https://doi.org/10.1521/soco.1988.6.4.287>
- Kédia, G., Berthoz, S., Wessa, M., Hilton, D., & Martinot, J.-L. (2008). An Agent Harms a Victim: A Functional Magnetic Resonance Imaging Study on Specific Moral Emotions. *Journal of Cognitive Neuroscience*, 20(10), 1788-1798. <https://doi.org/10.1162/jocn.2008.20070>
- Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation*, 15, 192-238.
- Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28(2), 107-128. <https://doi.org/10.1037/h0034225>

- Kelter, R. (2020). Bayesian alternatives to null hypothesis significance testing in biomedical research: A non-technical introduction to Bayesian inference with JASP. *BMC Medical Research Methodology*, 20(1), 142. <https://doi.org/10.1186/s12874-020-00980-6>
- Kerr, N. L., & Tindale, R. S. (2004). Group Performance and Decision Making. *Annual Review of Psychology*, 55(1), 623-655. <https://doi.org/10.1146/annurev.psych.55.090902.142009>
- King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9(2), 137-163. <https://doi.org/10.1093/oxfordjournals.pan.a004868>
- Kneer, M., & Bourgeois-Gironde, S. (2017). Mens rea ascription, expertise and outcome effects: Professional judges surveyed. *Cognition*, 169, 139-146. <https://doi.org/10.1016/j.cognition.2017.08.008>
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712-719. <https://doi.org/10.1016/j.tins.2004.10.007>
- Knobe, J., & Fraser, B. (2008). Causal Judgment and Moral Judgment: Two Experiments. Dans W. Sinnott-Armstrong (Ed.), *Moral Psychology*. MIT Press.
- Kogan, N., & Wallach, M. A. (1966). Modification of a judgmental style through group interaction. *Journal of Personality and Social Psychology*, 4(2), 165-174. <https://doi.org/10.1037/h0023566>
- Kominsky, J. F., & Phillips, J. (2019). Immoral Professors and Malfunctioning Tools: Counterfactual Relevance Accounts Explain the Effect of Norm Violations on Causal Selection. *Cognitive Science*, 43(11), Article 11. <https://doi.org/10.1111/cogs.12792>

- Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, *137*, 196-209. <https://doi.org/10.1016/j.cognition.2015.01.013>
- Korteling, J. E., & Toet, A. (2020). Cognitive Biases. Dans *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier. <https://doi.org/10.1016/B978-0-12-809324-5.24105-9>
- Kuhn, D. (2012). The development of causal reasoning. *WIREs Cognitive Science*, *3*(3), 327-335. <https://doi.org/10.1002/wcs.1160>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lagnado, D. (2011). Causal Thinking. Dans *Causality in the Sciences*. OUP Oxford.
- Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754-770. <https://doi.org/10.1016/j.cognition.2008.06.009>
- Lagnado, D. A., & Gerstenberg, T. (2017). Causation in Legal and Moral Reasoning. Dans M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 565–601). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199399550.013.30>
- Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal Responsibility and Counterfactuals. *Cognitive Science*, *37*(6), 1036-1073. <https://doi.org/10.1111/cogs.12054>
- Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, *8*(1), 33267. <https://doi.org/10.1525/collabra.33267>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159-174. <https://doi.org/10.2307/2529310>

- Larrick, R. P. (2004). Debiasing. *Blackwell handbook of judgment and decision making* (pp. 316-337). Blackwell Publishing. <https://doi.org/10.1002/9780470752937.ch16>
- Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22(3), 177-189. [https://doi.org/10.1016/0022-1031\(86\)90022-3](https://doi.org/10.1016/0022-1031(86)90022-3)
- Le Floch, V. (2008). Plausibility judgment of causally related social events: Impact of conceptual coherence and word-coherence. *Journal of Pragmatics*, 40(7), 1202-1215. <https://doi.org/10.1016/j.pragma.2008.01.002>
- Le Guen, O., Samland, J., Friedrich, T., Hanus, D., & Brown, P. (2015). Making sense of (exceptional) causal relations. A cross-cultural and cross-linguistic study. *Frontiers in Psychology*, 6, 1645. <https://doi.org/10.3389/fpsyg.2015.01645>
- Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility. *Personality and Social Psychology Bulletin*, 24(6), 563-574. <https://doi.org/10.1177/0146167298246001>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255. <https://doi.org/10.1037/0033-2909.125.2.255>
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511813313>
- Lewis, D. K. (1973). Causation. *Journal of Philosophy*, 70(17), 556-567. <https://doi.org/10.2307/2025310>
- Lewis, D. K. (2000). Causation as Influence. *Journal of Philosophy*, 97(4), 182-197. <https://doi.org/jphil200497437>

- Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 4(4), 390-398. <https://doi.org/10.1111/j.1745-6924.2009.01144.x>
- Lim, W., & Lim, A. (2016). A Comparison Of Usual t-Test Statistic and Modified t-Test Statistics on Skewed Distribution Functions. *Journal of Modern Applied Statistical Methods*, 15(2), 8. <https://doi.org/10.22237/jmasm/1478001960>
- Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, 358(6366), 1038-1041. <https://doi.org/10.1126/science.aag2132>
- Livengood, J., & Machery, E. (2007). The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence. *Midwest Studies in Philosophy*, 31, 107-127. <https://doi.org/10.1111/j.1475-4975.2007.00150.x>
- Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, 61(4), 303-332. <https://doi.org/10.1016/j.cogpsych.2010.05.002>
- 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>
- Lorenz, E. N. (1972, décembre). *Predictability: Does the Flap of a Butterfly's Wings in Brazil Set off a Tornado in Texas?* [Conférence]. American Association for the Advancement of Science. University of Washington.

- Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*(4), 955-984. <https://doi.org/10.1037/a0013256>
- Ludolph, R., & Schulz, P. J. (2018). Debiasing health-related judgments and decision making: A systematic review. *Medical Decision Making*, *38*(1), 3-13. <https://doi.org/10.1177/0272989X17716672>
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1-25. <https://doi.org/10.1086/392759>
- Mackie, J. L. (1980). *The Cement of the Universe: A Study of Causation*. Oxford University Press. <https://doi.org/10.1093/0198246420.001.0001>
- Malherbe, M. (1994). *Qu'est-ce que la causalité ? : Hume et Kant*. Vrin.
- Malle, B. F. (2011). Time to Give Up the Dogmas of Attribution: An Alternative Theory of Behavior Explanation. Dans J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (pp. 297-352). Academic Press. <https://doi.org/10.1016/B978-0-12-385522-0.00006-8>
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A Theory of Blame. *Psychological Inquiry*, *25*(2), 147-186. <https://doi.org/10.1080/1047840X.2014.877340>
- Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*(4), 661. <https://doi.org/10.1037/a0026790>
- Mandel, D. R. (2003). Judgment dissociation theory: An analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General*, *132*(3), 419. <https://doi.org/10.1037/0096-3445.132.3.419>

- Mandel, D. R. (2005). Counterfactual and causal explanation: From early theoretical views to new frontiers. Dans In D. R. Mandel, D. J. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 11-27). Routledge.
- Mandel, D. R. (2011). Mental Simulation and the Nexus of Causal and Counterfactual Explanation. Dans C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding Counterfactuals, Understanding Causation: Issues in Philosophy and Psychology* (pp. 147–170). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199590698.003.0008>
- Markman, K. D., & Tetlock, P. E. (2000). ‘I couldn’t have known’: Accountability, foreseeability and counterfactual denials of responsibility. *British Journal of Social Psychology*, 39(3), 313-325. <https://doi.org/10.1348/014466600164499>
- McCloy, R., & Byrne, R. M. J. (2000). Counterfactual thinking about controllable events. *Memory & Cognition*, 28(6), 1071-1078. <https://doi.org/10.3758/BF03209355>
- McClure, J., Hilton, D. J., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European Journal of Social Psychology*, 37(5), 879-901. <https://doi.org/10.1002/ejsp.394>
- McGill, A. L., & Tenbrunsel, A. E. (2000). Mutability and propensity in causal selection. *Journal of Personality and Social Psychology*, 79(5), 677-689. <https://doi.org/10.1037//0022-3514.79.5.677>
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57-74. <https://doi.org/10.1017/S0140525X10000968>

- Mill, J. S. (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139149839>
- Mischel, T. (1970). Human Action, Conceptual and Empirical Issues. *British Journal for the Philosophy of Science*, 21(1), 117-119.
- Mitchell, O. (2005). A Meta-Analysis of Race and Sentencing Research: Explaining the Inconsistencies. *Journal of Quantitative Criminology*, 21(4), 439-466. <https://doi.org/10.1007/s10940-005-7362-7>
- Monroe, A. E., & Malle, B. F. (2019). People systematically update moral judgments of blame. *Journal of Personality and Social Psychology*, 116(2), 215-236. <https://doi.org/10.1037/pspa0000137>
- Moore, M. S. (2009). The Role of Counterfactual Dependence as an Independent, Non-causal Desert-determiner. Dans M. S. Moore (Ed.), *Causation and Responsibility: An Essay in Law, Morals, and Metaphysics* (pp. 426–470). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199256860.003.0018>
- Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 129-140. <https://doi.org/10.1177/2372732215600886>
- Moscovici, S. (1972). *Introduction à la psychologie sociale*. Larousse.
- Moshman, D., & Geil, M. (1998). Collaborative Reasoning: Evidence for Collective Rationality. *Thinking & Reasoning*, 4(3), 231-248. <https://doi.org/10.1080/135467898394148>

- Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous Goal Inference (SGI). *Social and Personality Psychology Compass*, 10(1), 64-80.
<https://doi.org/10.1111/spc3.12232>
- Moutier, S., Angeard, N., & Houde, O. (2002). Deductive reasoning and matching-bias inhibition training: Evidence from a debiasing paradigm. *Thinking & Reasoning*, 8(3), 205-224. <https://doi.org/10.1080/13546780244000033>
- Moutier, S., & Houdé, O. (2003). Judgement under uncertainty and conjunction fallacy inhibition training. *Thinking & Reasoning*, 9(3), 185-201.
<https://doi.org/10.1080/13546780343000213>
- Myers, D. G., & Kaplan, M. F. (1976). Group-Induced Polarization in Simulated Juries. *Personality and Social Psychology Bulletin*, 2(1), 63-66.
<https://doi.org/10.1177/014616727600200114>
- Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations*, 9(2), 203-219.
<https://doi.org/10.1080/13869790600641905>
- Nadler, J. (2012). Blaming as a social process: the influence of character and moral emotion on blame. *Law and Contemporary Problems*, 75(2), 1-31.
- Nasri, A. (2006). *Contribution à l'étude de la notion de causalité en droit pénal* [Thèse de doctorat, Paris 1]. <https://www.theses.fr/2006PA010289>
- Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences*, 107(40), 17140-17145. <https://doi.org/10.1073/pnas.0914056107>

- Neys, W. D. (2018). *Dual Process Theory 2.0*. Routledge.
<https://doi.org/10.4324/9781315204550>
- Nisbett, R. E., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Prentice Hall Inc. <https://philpapers.org/rec/nishis>
- Open science collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Operskalski, J. T., & Barbey, A. K. (2017). Cognitive Neuroscience of Causal Reasoning. In M. R. Waldmann (Ed.), *The Oxford Handbook of Causal Reasoning* (pp. 217–242). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199399550.013.16>
- Paul, L. A., Hall, N., & Hall, E. J. (2013). *Causation: A User's Guide*. Oxford University Press.
- Pearl, J. (1998). *On the Definition of Actual Cause*. Technical Report R-259, Department of Computer Science, University of California, Los Angeles, Calif.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.
- Peer, E., & Gamliel, E. (2013). Heuristics and Biases in Judicial Decisions. *Court Review*, 49, 114.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, 16(10), 1706-1716. <https://doi.org/10.1162/0898929042947900>
- Pennequin, V. (2021). Métacognition et flexibilité : Quels liens théoriques et quels liens observés ? Dans *La flexibilité cognitive : Pierre angulaire de l'apprentissage* (pp. 59-75). ISTE Editions. <http://dx.doi.org/10.51926/iste.9035.ch3>

- Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*(2), 242-258. <https://doi.org/10.1037/0022-3514.51.2.242>
- Perales, J. C., Catena, A., Cándido, A., & Maldonado, A. (2017). Rules of causal judgment: Mapping statistical information onto causal beliefs. Dans *The Oxford handbook of causal reasoning* (pp. 29-51). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199399550.001.0001>
- Petrocelli, J. V., Percy, E. J., Sherman, S. J., & Tormala, Z. L. (2011). Counterfactual potency. *Journal of Personality and Social Psychology*, *100*(1), 30-46. <https://doi.org/10.1037/a0021523>
- Pettigrove, G., & Tanaka, K. (2014). Anger and moral judgment. *Australasian Journal of Philosophy*, *92*(2), 269-286. <https://doi.org/10.1080/00048402.2013.795990>
- Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, *114*(18), 4649-4654. <https://doi.org/10.1073/pnas.1619717114>
- Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, *145*, 30-42. <https://doi.org/10.1016/j.cognition.2015.08.001>
- Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026-1040. <https://doi.org/10.1016/j.tics.2019.09.007>
- Phillips, J., & Shaw, A. (2015). Manipulating Morality: Third-Party Intentions Alter Moral Judgments by Changing Causal Reasoning. *Cognitive Science*, *39*(6), 1320-1347. <https://doi.org/10.1111/cogs.12194>

- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4), 717-731. <https://doi.org/10.3758/BF03206553>
- Quatto, P., Ripamonti, E., & Marasini, D. (2020). Best uses of p-values and complementary measures in medical research: Recent developments in the frequentist and Bayesian frameworks. *Journal of Biopharmaceutical Statistics*, 30(1), 121-142. <https://doi.org/10.1080/10543406.2019.1632874>
- Quézel-Ambrunaz, C. (2008). *Essai sur la causalité en droit de la responsabilité civile* [Thèse de doctorat, Chambéry]. <https://www.theses.fr/2008CHAML010>
- Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry*, 18(1), 1-8. <https://doi.org/10.1186/s12888-018-1761-4>
- Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, 118(1), 57-75. <https://doi.org/10.1037/a0021867>
- Reuter, K., Kirfel, L., van Riel, R., & Barlassina, L. (2014). The good, the bad, and the timely: How temporal order and moral judgment influence causal selection. *Frontiers in Psychology*, 5, 1-10. <https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01336>
- Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, Development, and Rationality: An Integrative Theory of Judgment and Decision Making. Dans J. Shanteau & S. L. Schneider (Éds.), *Emerging Perspectives on Judgment and Decision Research* (p. 201-245). Cambridge University Press. <https://doi.org/10.1017/CBO9780511609978.009>

- Robbennolt, J. K. (2000). Outcome Severity and Judgments of “Responsibility”: A Meta-Analytic Review. *Journal of Applied Social Psychology, 30*(12), 2575-2609. <https://doi.org/10.1111/j.1559-1816.2000.tb02451.x>
- Rogers, R., Alicke, M. D., Taylor, S. G., Rose, D., Davis, T. L., & Bloom, D. (2019). Causal deviance and the ascription of intent and blame. *Philosophical Psychology, 32*(3), 404-427. <https://doi.org/10.1080/09515089.2018.1564025>
- Rose, D., Sievers, E., & Nichols, S. (2021). Cause and Burn. *Cognition, 207*, 104517. <https://doi.org/10.1016/j.cognition.2020.104517>
- Ross, L. (1977). The intuitive psychologist and his shortcomings: distortions in the attribution process. Dans *Advances in experimental social psychology* (pp. 173-220). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60357-3](https://doi.org/10.1016/S0065-2601(08)60357-3)
- Rosset, E. (2008). It’s no accident: Our bias for intentional explanations. *Cognition, 108*(3), 771-780. <https://doi.org/10.1016/j.cognition.2008.07.001>
- Rotter, J. B. (1954). *Social learning and clinical psychology*. Prentice-Hall, Inc. <https://doi.org/10.1037/10788-000>
- Rottman, B. M. (2017). The acquisition and use of causal structure knowledge. Dans *The Oxford handbook of causal reasoning* (pp. 85-114). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199399550.001.0001>
- Russell, P. S., & Giner-Sorolla, R. (2011). Moral anger, but not moral disgust, responds to intentionality. *Emotion, 11*(2), 233-240. <https://doi.org/10.1037/a0022598>
- Saladin, M., Saper, Z., & Breen, L. (1988). Perceived attractiveness and attributions of criminality: What is beautiful is not criminal. *Canadian Journal of Criminology, 30*(3), 251-259.

- Samland, J., Proft, M., & Waldmann, M. (2016). The Role of Prescriptive Norms and Knowledge in Children's and Adults' Causal Selection. *Journal of Experimental Psychology: General*, *145*(2), 125-130.
- Samland, J., & Waldmann, M. (2015). Highlighting the Causal Meaning of Causal Test Questions in Contexts of Norm Violations. *Cognitive Science*.
<https://www.semanticscholar.org/paper/Highlighting-the-Causal-Meaning-of-Causal-Test-in-Samland-Waldmann/7cff09d7e048294ba2638a5264b1218621a1f3ca>
- Samland, J., & Waldmann, M. R. (2014). Do Social Norms Influence Causal Inferences? Dans *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36).
<https://escholarship.org/uc/item/6qc3p3mb>
- Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, *156*, 164-176. <https://doi.org/10.1016/j.cognition.2016.07.007>
- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144-1167. <https://doi.org/10.1037/a0020511>
- Sarin, A., Lagnado, D. A., & Burgess, P. W. (2017). The Intention-Outcome Asymmetry Effect. *Experimental Psychology*, *64*(2), 124-141. <https://doi.org/10.1027/1618-3169/a000359>
- Satpute, A. B., Fenker, D. B., Waldmann, M. R., Tabibnia, G., Holyoak, K. J., & Lieberman, M. D. (2005). An fMRI study of causal judgments. *European Journal of Neuroscience*, *22*(5), 1233-1238. <https://doi.org/10.1111/j.1460-9568.2005.04292.x>

- Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret Agents: Inferences About Hidden Causes by 10- and 12-Month-Old Infants. *Psychological Science*, *16*(12), 995-1001. <https://doi.org/10.1111/j.1467-9280.2005.01649.x>
- Schaffer, J. (2005). Contrastive Causation. *The Philosophical Review*, *114*(3), 327-358. <https://www.jstor.org/stable/30043679>
- Schank, R. C., & Abelson, R. P. (1988). *Scripts, plans, goals, and understanding*. Psychology Press.
- Schleifer, M., Shultz, T. R., & Lefebvre-Pinard, M. (1983). Children's judgements of causality, responsibility and punishment in cases of harm due to omission. *British Journal of Developmental Psychology*, *1*(1), 87-97. <https://doi.org/10.1111/j.2044-835X.1983.tb00546.x>
- Schwarz, N. (1995). What respondents learn from questionnaires: The survey interview and the logic of conversation. *International Statistical Review / Revue Internationale de Statistique*, *63*(2), 153-168. <https://doi.org/10.2307/1403610>
- Schwarz, N., Strack, F., Hilton, D.J., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition*, *9*(1), 67-84. <https://doi.org/10.1521/soco.1991.9.1.67>
- Sedlmeier, P., & Hilton, D. J. (2012). Improving judgment and decision making through communication and representation. Dans *Judgment and decision making as a skill: Learning, development and evolution* (pp. 229-257). Cambridge University Press.
- Sellier, A.-L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing Training Improves Decision Making in the Field. *Psychological Science*, *30*(9), 1371-1379. <https://doi.org/10.1177/0956797619861429>

- Sequeira, H., & D'Hondt, F. (2013). Chapitre 6. L'activité électrodermale dans l'étude de la cognition. Dans *Electrophysiologie de la cognition* (pp. 137-156). Dunod.
<https://doi.org/10.3917/dunod.hot.2013.01.0137>
- Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, 53(1), 491-517.
<https://doi.org/10.1146/annurev.psych.53.100901.135213>
- Shaver, K. G. (1985). Attributions of Responsibility. Dans K. G. Shaver (Éd.), *The Attribution of Blame: Causality, Responsibility, and Blameworthiness* (pp. 87-113). Springer.
https://doi.org/10.1007/978-1-4612-5094-4_5
- Shultz, T. R., Schleifer, M., & Altman, I. (1981). Judgments of causation, responsibility, and punishment in cases of harm-doing. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 13(3), 238–253.
<https://doi.org/10.1037/h0081183>
- Shultz, T. R., & Wright, K. (1985). Concepts of negligence and intention in the assignment of moral responsibility. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement*, 17(2), 97–108. <https://doi.org/10.1037/h0080138>
- Shultz, T. R., Wright, K., & Schleifer, M. (1986). Assignment of Moral Responsibility and Punishment. *Child Development*, 57(1), 177-184. JSTOR.
<https://doi.org/10.2307/1130649>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359-1366.
<https://doi.org/10.1177/0956797611417632>

- Simon, H. A. (1990). Bounded Rationality. Dans J. Eatwell, M. Milgate, & P. Newman (Éds.), *Utility and Probability* (pp. 15-18). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-20568-4_5
- Sloman, S., Barbey, A. K., & Hotaling, J. M. (2009). A Causal Model Theory of the Meaning of Cause, Enable, and Prevent. *Cognitive Science*, 33(1), 21-50. <https://doi.org/10.1111/j.1551-6709.2008.01002.x>
- Slugoski, B. R., Lalljee, M., Lamb, R., & Ginsburg, G. P. (1993). Attribution in conversational context: Effect of mutual knowledge on explanation-giving. *European Journal of Social Psychology*, 23(3), 219-238. <https://doi.org/10.1002/ejsp.2420230302>
- Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A User's Guide to Debiasing. Dans *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 924-951). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118468333.ch33>
- Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, 126(4), 323-348. <https://doi.org/10.1037/0096-3445.126.4.323>
- Spellman, B. A., & Gilbert, E. A. (2014). Blame, Cause, and Counterfactuals: The Inextricable Link. *Psychological Inquiry*, 25(2), 245-250. <https://doi.org/10.1080/1047840X.2014.901708>
- Sperber, D., & Wilson, D. (1996). *Relevance: Communication and Cognition*, (2nd Edition). Wiley.
- Stanovich, K. E. (2018). Miserliness in human cognition: The interaction of detection, override and mindware. *Thinking & Reasoning*, 24(4), 423-444. <https://doi.org/10.1080/13546783.2018.1459314>

- Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The Rationality Quotient: Toward a Test of Rational Thinking*. MIT Press.
- Stapleton, J. (1988). Law, Causation and Common Sense. *Oxford Journal of Legal Studies*, 8(1), 111-131.
- Stapleton, J. (2008). Choosing What We Mean by Causation in the Law. *Missouri Law Review*, 73(2), 433.
- Stapleton, J. (2015). An ‘Extended But-For’ Test for the Causal Relation in the Law of Obligations. *Oxford Journal of Legal Studies*, 35(4), 697-726. <https://doi.org/10.1093/ojls/gqv005>
- Strickland, B., Silver, I., & Keil, F. C. (2017). The texture of causal construals : Domain-specific biases shape causal inferences from discourse. *Memory & Cognition*, 45(3), 442–455. <https://doi.org/10.3758/s13421-016-0668-x>
- Stiegler, B. (2022, 7 avril). *L'idéologie des biais cognitifs* [Conférence]. La Fabrique du citoyen 2022, Bibliothèque de Bordeaux. https://www.youtube.com/watch?v=Z71oV00aqxk&ab_channel=Biblioth%C3%A8quesdeBordeaux
- Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541-542. <https://doi.org/10.4300/JGME-5-4-18>
- Summers, A. (2018). Common-Sense Causation in the Law. *Oxford Journal of Legal Studies*, 38(4), 793-821. <https://doi.org/10.1093/ojls/gqy028>
- Sunstein, C. R. (2002). The Law of Group Polarization. *Journal of Political Philosophy*, 10(2), 175-195. <https://doi.org/10.1111/1467-9760.00148>

- Symborski, C., Barton, M., Quinn, M., Morewedge, C. K., Kassam, K., & Korris, J. (2014). *Missing: A Serious Game for the Mitigation of Cognitive Biases*. Dans Interservice/Industry Training, Simulation, and Education Conference (IITSEC) (pp. 1-13).
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, *109*(3), 451-471. <https://doi.org/10.1037/0033-295X.109.3.451>
- Tetlock, P. E., Lerner, J. S., & Boettger, R. (1996). The dilution effect: Judgmental bias, conversational convention, or a bit of both? *European Journal of Social Psychology*, *26*(6), 915-934. [https://doi.org/10.1002/\(SICI\)1099-0992\(199611\)26:6<915::AID-EJSP797>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1099-0992(199611)26:6<915::AID-EJSP797>3.0.CO;2-W)
- Tetlock, P. E., & Mellers, B. A. (2002). The Great Rationality Debate. *Psychological Science*, *13*(1), 94-99. <https://doi.org/10.1111/1467-9280.00418>
- Thomas, E. A. C., & Parpal, M. (1987). Liability as a function of plaintiff and defendant fault. *Journal of Personality and Social Psychology*, *53*(5), 843-857. <https://doi.org/10.1037/0022-3514.53.5.843>
- Thompson, V. A., & Markovits, H. (2021). Reasoning strategy vs cognitive capacity as predictors of individual differences in reasoning performance. *Cognition*, *217*, 532–543. <https://doi.org/10.1016/j.cognition.2021.104866>
- Thompson, V. A., Pennycook, G., Trippas, D., & Evans, J. St. B. T. (2018). Do smart people have better intuitions? *Journal of Experimental Psychology: General*, *147*(7), 945-961. <https://doi.org/10.1037/xge0000457>

- Tobia, K. (2020). Legal Concepts and Legal Expertise. *Synthese*.
<https://doi.org/10.2139/ssrn.3536564>
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2017). Real-World Correlates of Performance on Heuristics and Biases Tasks in a Community Sample. *Journal of Behavioral Decision Making*, 30(2), 41-554. <https://doi.org/10.1002/bdm.1973>
- Trabasso, T., & Bartolone, J. (2003). Story understanding and counterfactual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 904–923. <https://doi.org/10.1037/0278-7393.29.5.904>
- Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, 24(5), 595-611. [https://doi.org/10.1016/0749-596X\(85\)90048-8](https://doi.org/10.1016/0749-596X(85)90048-8)
- Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, 24(5), 612-630. [https://doi.org/10.1016/0749-596X\(85\)90049-X](https://doi.org/10.1016/0749-596X(85)90049-X)
- Trémolière, B., & Neys, W. D. (2013). Methodological concerns in moral judgement research: Severity of harm shapes moral decisions. *Journal of Cognitive Psychology*, 25(8), 989-993. <https://doi.org/10.1080/20445911.2013.841169>
- Turri, J., & Blouw, P. (2015). Excuse validation: A study in rule-breaking. *Philosophical Studies*, 172(3), 615-634. <https://doi.org/10.1007/s11098-014-0322-z>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131. <https://doi.org/10.1126/science.185.4157.1124>
- Tversky, A., & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *The Journal of Business*, 59(4), 251-278.

- Umpress, E. E., Simmons, A. L., Folger, R., Ren, R., & Bobocel, R. (2013). Observer reactions to interpersonal injustice: The roles of perpetrator intent and victim perception. *Journal of Organizational Behavior*, 34(3), 327-349. <https://doi.org/10.1002/job.1801>
- Usher, M. (2020). Agency, Teleological Control and Robust Causation. *Philosophy and Phenomenological Research*, 100(2), 302-324. <https://doi.org/10.1111/phpr.12537>
- Valdesolo, P., & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44(5), 1334-1338. <https://doi.org/10.1016/j.jesp.2008.03.010>
- van den Bergh, D., van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., Dablander, F., Gronau, Q. F., Kucharský, Š., Gupta, A. R. K. N., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.-J. (2020). A Tutorial on Conducting and Interpreting a Bayesian ANOVA in JASP. *L'Année psychologique*, 120(1), 73-96. <https://doi.org/10.3917/anpsy1.201.0073>
- Van der Henst, J.-B. (2002). La perspective pragmatique dans l'étude du raisonnement et de la rationalité. *L'Année psychologique*, 102(1), 65-108. <https://doi.org/10.3406/psy.2002.29583>
- Van Hoeck, N., Watson, P. D., & Barbey, A. K. (2015). Cognitive neuroscience of human counterfactual reasoning. *Frontiers in Human Neuroscience*, 9, 420. <https://doi.org/10.3389/fnhum.2015.00420>
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829-858. <https://doi.org/10.1002/hbm.20547>

- Vasilyeva, N., Blanchard, T., & Lombrozo, T. (2018). Stable causal relationships are better causal relationships. *Cognitive Science*, 42(4), 1265-1296. <https://doi.org/10.1111/cogs.12605>
- Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, 1224(1), 22-39. <https://doi.org/10.1111/j.1749-6632.2011.05965.x>
- Vinokur, A., & Burnstein, E. (1978). Depolarization of attitudes in groups. *Journal of Personality and Social Psychology*, 36(8), 872-885. <https://doi.org/10.1037/0022-3514.36.8.872>
- Voiklis, J., & Malle, B. F. (2018). Moral cognition and its basis in social cognition and social regulation. Dans *Atlas of moral psychology* (pp. 108-120). The Guilford Press.
- Von Helversen, B., & Rieskamp, J. (2009). Predicting sentencing for low-level crimes: Comparing models of human judgment. *Journal of Experimental Psychology. Applied*, 15(4), 375-395. <https://doi.org/10.1037/a0018024>
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35-57. <https://doi.org/10.3758/s13423-017-1343-3>
- Wagner-Egger, P. (2011). Les canons de la rationalité : Essai de classification des points de vue dans le débat sur les biais cognitifs et la rationalité humaine. *L'Année psychologique*, 111(1), 191-224. <https://doi.org/10.3917/anpsy.111.0191>
- Waldmann, M. (2017). *The Oxford Handbook of Causal Reasoning*. Oxford University Press.

- Waldmann, M., & Hagmayer, Y. (2013). Causal reasoning. Dans *Oxford Handbook of Cognitive Psychology* (pp. 733-752). Oxford Library of Psychology.
- Walsh, C. R., & Byrne, R. M. J. (2007). The Effects of Reasons for Acting on Counterfactual Thinking. *Thinking and Reasoning*, 13, 461-483.
- Walsh, C. R., & Sloman, S. A. (2011). The Meaning of Cause and Prevent: The Role of Causal Mechanism. *Mind & Language*, 26(1), 21-52. <https://doi.org/10.1111/j.1468-0017.2010.01409.x>
- Wason, P. C. (1960). On the Failure to Eliminate Hypotheses in a Conceptual Task. *Quarterly Journal of Experimental Psychology*, 12(3), 129-140. <https://doi.org/10.1080/17470216008416717>
- Weiner, B. (1985). « Spontaneous » causal thinking. *Psychological Bulletin*, 97(1), 74-84. <https://doi.org/10.1037/0033-2909.97.1.74>
- Weiner, B., Amirkhan, J., Folkes, V. S., & Verette, J. A. (1987). An attributional analysis of excuse giving: Studies of a naive theory of emotion. *Journal of Personality and Social Psychology*, 52(2), 316-324. <https://doi.org/10.1037/0022-3514.52.2.316>
- Wente, A., Gopnik, A., Fernández Flecha, M., Garcia, T., & Buchsbaum, D. (2022). Causal learning, counterfactual reasoning and pretend play: A cross-cultural comparison of Peruvian, mixed- and low-socioeconomic status U.S. children. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1866), 20210345. <https://doi.org/10.1098/rstb.2021.0345>
- White, A. R. (1968). *The Philosophy of Action*. London: Oxford University Press.
- White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin*, 108(1), 3-18. <https://doi.org/10.1037/0033-2909.108.1.3>

- Willemsen, P., & Kirfel, L. (2019). Recent empirical work on the relationship between causal judgements and norms. *Philosophy Compass*, *14*(1), e12562. <https://doi.org/10.1111/phc3.12562>
- Willingham, D. T. (2008). Critical Thinking: Why Is It So Hard to Teach? *Arts Education Policy Review*, *109*(4), 21-32. <https://doi.org/10.3200/AEPR.109.4.21-32>
- Willmott, D., Boduszek, D., Debowska, A., & Woodfield, R. (2018). Introduction and validation of the Juror Decision Scale (JDS): An empirical investigation of the Story Model. *Journal of Criminal Justice*, *57*, 26-34. <https://doi.org/10.1016/j.jcrimjus.2018.03.004>
- Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology. General*, *136*(1), 82–111. <https://doi.org/10.1037/0096-3445.136.1.82>
- Wong, P. T., & Weiner, B. (1981). When people ask « why » questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, *40*(4), 650-663. <https://doi.org/10.1037/0022-3514.40.4.650>
- Woodward, J. (2006). Sensitive and Insensitive Causation. *The Philosophical Review*, *115*(1), 1-50.
- Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, *25*(3), 287-318. <https://doi.org/10.1007/s10539-010-9200-z>
- Wright, R. W. (1985). Causation in Tort Law. *California Law Review*, *73*, 1735.
- Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148), 1075-1080. <https://doi.org/10.1038/nature05852>

Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, 37(2), 197-206.
<https://doi.org/10.1086/651257>

Annexes

Index des Annexes

Annexe A - Exemple de questionnaire de l'Étude 1	II
Annexe B - Exemple de questionnaire de l'Étude 2	III
Annexe C - Scores d'évaluation de l'Étude 2	IV
Annexe D - Scénarios des Études 3 et 4	V
Annexe E - Études de la revue narrative de la littérature	VII
Annexe F – Scores d'évaluation de l'Étude 5	X

Annexe A

Exemple de questionnaire de l'Étude 1

Veillez d'abord indiquer :

Votre genre : F () H ()

Votre âge : _____

Veillez lire attentivement le récit ci-dessous, avant de répondre aux questions posées. Merci de votre participation !

Barbara est sévèrement et continuellement battue par son mari, Jean, et veut le tuer. Au cours d'un repas au restaurant, elle envisage de verser du poison dans le plat de Jean à son insu.

A peine après avoir commencé son plat, Jean s'absente un instant pour se rendre aux toilettes.

Durant son absence, elle verse le poison dans le plat de Jean, un poison aux effets très rapides.

A son retour, Jean poursuit son repas, et décède dans les minutes qui suivent les premières bouchées.

Question 1 : Quelle est la cause réelle de la mort de Jean ?

Question 2 : Envers qui ou quoi faites-vous le plus de reproches ?

Annexe B

Exemple de questionnaire de l'Étude 2

Veillez d'abord indiquer :

Votre genre : F () H ()

Votre âge : _____

Veillez lire attentivement le récit ci-dessous, avant de répondre aux questions posées. Merci de votre participation !

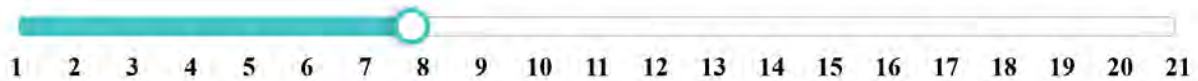
Barbara est sévèrement et continuellement battue par son mari, Jean, et veut le tuer.

Au cours d'un repas au restaurant, elle envisage de verser du poison dans le plat de Jean à son insu. À peine avoir commencé son plat, Jean s'absente un instant pour se rendre aux toilettes.

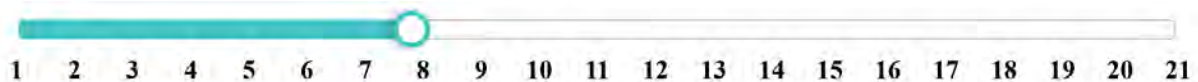
Durant son absence, elle verse le poison dans le plat de Jean, un poison aux effets très rapides.

A son retour, Jean poursuit son repas, et décède dans les minutes qui suivent les premières bouchées.

Question 1 : Comment évalueriez-vous la décision de Barbara de vouloir tuer Jean ? (1 = Pas du tout blâmable à 21 = Tout à fait blâmable)



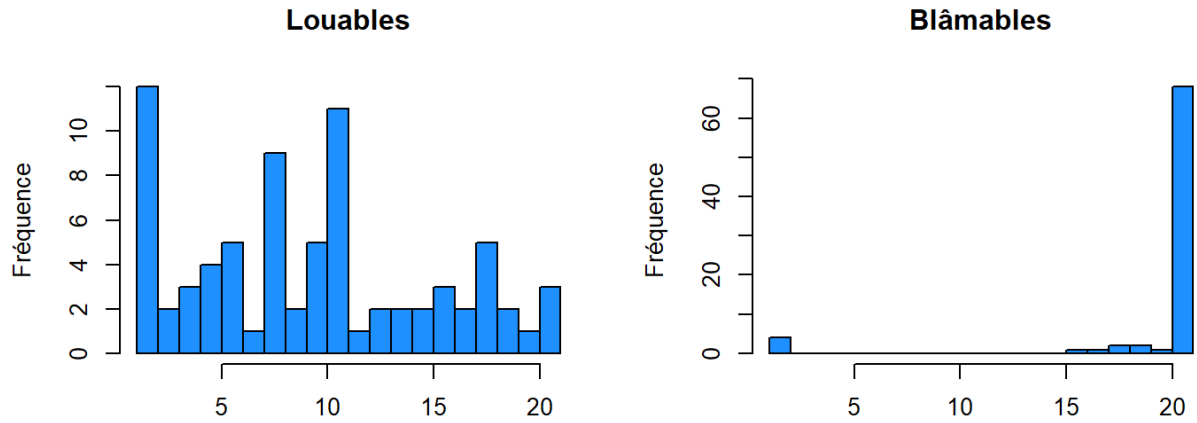
Question 2 : Dans quelle mesure les actions et les décisions de Barbara sont-elles la cause de la mort de Jean ? (1 = Pas du tout la cause à 21 = Tout à fait la cause)



Annexe C

Figure C1

Distributions des scores d'évaluation en fonction des motivations de l'agent dans l'Étude 2.



Note. Echelle allant de 1 (pas du tout blâmable) à 21 (totalement blâmable)

Résultat du test de Levene : $F(1, 154) = 20.6, p < .001$ (hétérogénéité des variances).

Annexe D

Scénarios des Études 3 et 4

Scénario du Feu de forêt

Lien causal présent /agent louable

Un village forestier était encerclé par un feu de forêt et était complètement coupé de toute aide extérieure. Cet incendie risquait d'atteindre le village. Apprenant la nouvelle, David, un garde-forestier, mit délibérément le feu à une zone de la forêt. Son but était de créer un contre-feu pour empêcher le premier d'atteindre le village. Il avait pris ses précautions en se renseignant sur les conditions météorologiques : aucun vent n'était prévu pour la journée. Son feu pouvait se faire en toute sécurité.

Au bout d'un moment, le premier feu s'éteignit et celui de David fut sur le point de faire de même lorsqu'une improbable légère brise vint le raviver. Le feu atteignit donc le village. Douze personnes y perdirent la vie.

Lien causal présent /agent blâmable

Un village forestier était encerclé par un feu de forêt et était complètement coupé de toute aide extérieure. Cet incendie risquait d'atteindre le village. Apprenant la nouvelle, David, un pyromane, mit délibérément le feu à une zone de la forêt. Son but était de créer un feu qui permettrait au premier d'atteindre plus facilement le village.

Au bout d'un moment, le premier feu s'éteignit et celui de David fut sur le point de faire de même lorsqu'une légère brise vint le raviver. Le feu atteignit donc le village et douze personnes perdirent la vie.

Lien causal absent /agent louable

Un village forestier était encerclé par un feu de forêt et était complètement coupé de toute aide extérieure. Cet incendie risquait d'atteindre le village. Apprenant la nouvelle, David, un garde-forestier, mit délibérément le feu à une zone de la forêt. Son but était de créer un contre-feu pour empêcher le premier d'atteindre le village. Il avait pris ses précautions en se renseignant sur les conditions météorologiques : aucun vent n'était prévu pour la journée. Son feu pouvait se faire en toute sécurité.

Au bout d'un moment, les deux feux s'éteignirent. Toutefois, la foudre vint frapper un arbre et un nouveau feu démarra. Une improbable légère brise vint le nourrir créant ainsi un grand incendie, qui atteignit le village. Douze personnes y perdirent la vie.

Lien causal absent /agent blâmable

Un village forestier était encerclé par un feu de forêt et était complètement coupé de toute aide extérieure. Cet incendie risquait d'atteindre le village. Apprenant la nouvelle, David, un pyromane, mit délibérément le feu à une zone de la forêt. Son but était de créer un feu qui permettrait au premier d'atteindre plus facilement le village.

Au bout d'un moment, les deux feux s'éteignirent. Toutefois, la foudre vint frapper un arbre et un nouveau feu démarra. Une légère brise vint le nourrir créant ainsi un grand incendie qui atteignit le village. Douze personnes y perdirent la vie.

Scénario de la Salmonelle

Lien causal présent /agent louable

Le ministère de la Santé annonça à tous les restaurants du pays qu'il y avait un risque accru de prolifération de la salmonelle en cette période. Dans un des restaurants, Quentin, un employé, souhaite diminuer la température de la chambre froide pour réduire les risques d'intoxication alimentaire par les clients. Malheureusement, le thermostat avait subi un dysfonctionnement. En voulant baisser la température, Quentin l'avait augmenté. Par conséquent, la chambre froide se réchauffa et les salmonelles purent proliférer. Le lendemain, plusieurs clients tombèrent malades et le restaurant dû fermer ses portes.

Lien causal présent /agent blâmable

Le ministère de la Santé annonça à tous les restaurants du pays qu'il y avait un risque accru de prolifération de la salmonelle en cette période. Dans un des restaurants, Quentin, un employé, souhaite augmenter la température de la chambre froide pour croître les risques d'intoxication alimentaire par les clients. Néanmoins, le thermostat avait subi un dysfonctionnement. En voulant augmenter la température, Quentin l'avait baissé. Ce refroidissement intense provoqua une surchauffe du système de refroidissement qui cessa de fonctionner. Par conséquent, la chambre froide se réchauffa et les salmonelles purent proliférer. Le lendemain, plusieurs clients tombèrent malades et le restaurant dû fermer ses portes.

Lien causal absent /agent louable

Le ministère de la Santé annonça à tous les restaurants du pays qu'il y avait un risque accru de prolifération de la salmonelle en cette période. Dans un des restaurants, Quentin, un employé, souhaite diminuer la température de la chambre froide pour réduire les risques d'intoxication alimentaire par les clients. Le thermostat avait subi un dysfonctionnement. La température n'avait donc pas été modifiée sous l'action de Quentin. En revanche, le dysfonctionnement provoqua un réchauffement de la chambre froide. Par conséquent, les salmonelles purent proliférer. Le lendemain, plusieurs clients tombèrent malades et le restaurant dû fermer ses portes.

Lien causal absent /agent blâmable

Le ministère de la Santé annonça à tous les restaurants du pays qu'il y avait un risque accru de prolifération de la salmonelle en cette période. Dans un des restaurants, Quentin, un employé, souhaite augmenter la température de la chambre froide pour croître les risques d'intoxication alimentaire par les clients. Le thermostat avait subi un dysfonctionnement. La température n'avait donc pas été modifiée sous l'action de Quentin. En revanche, le dysfonctionnement provoqua un réchauffement de la chambre froide. Par conséquent, les salmonelles purent proliférer. Le lendemain, plusieurs clients tombèrent malades et le restaurant dû fermer ses portes.

Annexe E

Tableau E1.

Synthèse des études de la revue narrative de la littérature sur l'effet de la valence morale

Référence	Étude	Caractéristiques des participants	N	Age moyen (SD)	Femmes / Hommes	Modèle causal	Type de question	Réplication des effets de BV
Alicke (1992)	1	étudiants en licence de psychologie	174	abs	abs	2	1 + 2	1
	3	étudiants en psychologie	112	abs	abs	3	2	1
	4	étudiants en psychologie	382	abs	abs	1 + 2 + 4	2	1
Alicke et al. (2008)	1	étudiants en licence de psychologie	102	abs	abs	1 + 3	2	1
	2	étudiants en licence de psychologie	270	abs	abs	2 + 3	2	1
	3	étudiants en psychologie	91	abs	abs	1 + 3	2	1
Alicke et al. (2011)	1	étudiants en licence de psychologie	319	abs	121/193	2	2	1
	2	abs	abs	abs	abs	2	2	1
	3	abs	abs	abs	abs	1 + 4	2	1 + 0
Clarke et al. (2015)	1	utilisateurs d'Amazon Mechanical Turk	587	abs	abs	2	2	1
	2	utilisateurs d'Amazon Mechanical Turk	197	abs	abs	2	2	1
Devos-Comby & Devos (2001)	1	lycéens suisses francophones	95	abs	57/37	2	2	0
Fincham & Shultz (1981)	1	étudiants d'une petite ville anglaise	60	abs	abs	2	2	1
Gilbert et al. (2015)	2	étudiants en psychologie de l'Université de Virginie	210	abs	abs	2	2	1
Hilton et al. (2016)	1	étudiants de l'Université Victoria de Wellington	111	abs	abs	2	2	1
	3	étudiants en psychologie à l'Université Victoria de Wellington	232	abs	abs	2	2	1
	4	étudiants en psychologie de l'Université Victoria de Wellington	192	abs	abs	2	2	1
Hilton et al. (2021)	3	français recrutés en ligne	206	25.95 (abs)	156/50	3 + 4	2	0

Icard et al. (2017)	1	utilisateurs d'Amazon Mechanical Turk (USA only)	426	abs	abs	2	2	1 + 0
Kanekar et al. (1993)	1	étudiants en licence de l'Université de Bombay	720	abs	abs	1	2	1
Kominsky & Phillips (2019)	1	utilisateurs d'Amazon Mechanical Turk	439	37.28 (12.14)	338 /101	2	2	1 + 0
	2	utilisateurs d'Amazon Mechanical Turk	603	37.03 (12.04)	271/329	2	2	1 + 0
	4	utilisateurs d'Amazon Mechanical Turk	403	34.96 (11.90)	197/205	2	2	1 + 0
Kominsky et al. (2015)	1	utilisateurs d'Amazon Mechanical Turk	60	abs	abs	2	2	1
	2	utilisateurs d'Amazon Mechanical Turk	120	abs	abs	2	2	1
Lagnado & Channon (2008)	1	étudiants en licence de l'University College of London	80	abs	abs	2	2	1
	2	étudiants en licence de l'University College of London	80	abs	abs	1 + 3	2	1
Nadler (2012)	1	utilisateurs d'Amazon Mechanical Turk	194	41 (abs)	118/76	1	2	0
Phillips et al. (2015)	2.a	utilisateurs d'Amazon Mechanical Turk	399	29.10 (8.84)	154/245	2	2	1
	3.a	utilisateurs de Amazon's Mechanical Turk	399	31.62 (10.50)	157/242	1	2	1
Phillips & Shaw (2015)	3	utilisateurs de Amazon's Mechanical Turk	125	abs	53/72	1	2	1
	4a	utilisateurs de Amazon's Mechanical Turk	193	30.77 (10.15)	66/127	1	2	1
	4b	utilisateurs de Amazon's Mechanical Turk	405	29.85 (10.40)	128/277	1	2	1
Rogers et al. (2019)	3	étudiants en licence américains	299	abs	abs	1	2	1
Samland & Waldmann (2014)	1	britanniques en ligne	206	abs	abs	2	2	1 + 0
	2	participants en ligne	99	abs	abs	2	2	1 + 0
Samland & Waldmann (2015)	1	étudiants de l'Université de Gottingen	72	abs	abs	2	2	1 + 0
	2	britanniques en ligne	98	abs	abs	2	1	1 + 0
Samland & Waldmann (2016)	1	participants recrutés en ligne de différents pays	134	30.01 (9.31)	abs	2	2	1 + 0

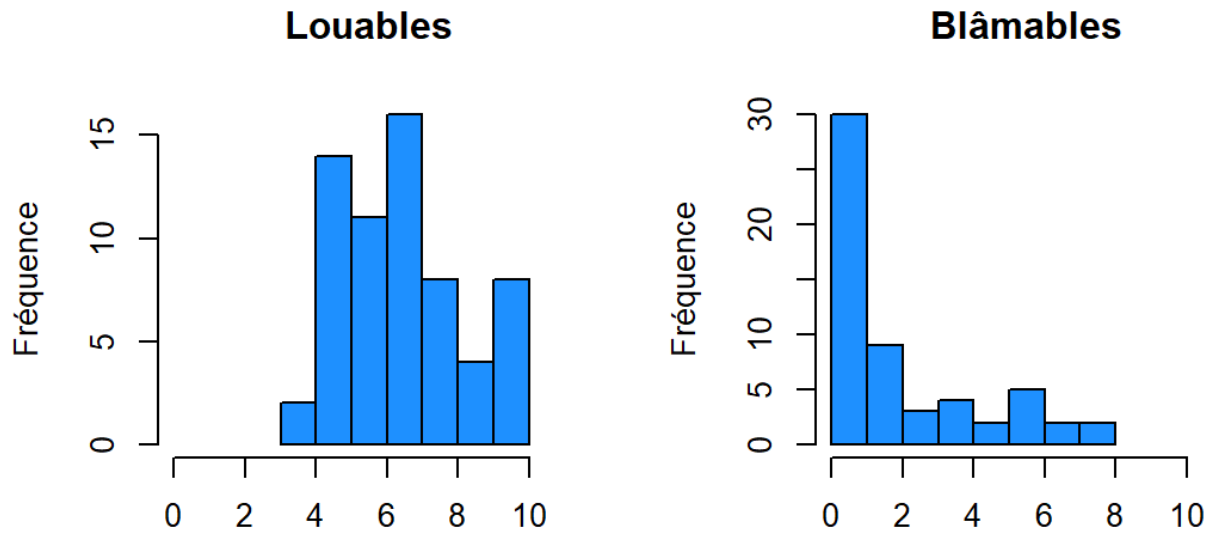
	2.a	britanniques recrutés en ligne	133	40.72 (12.77)	abs	2	2	1
	3	allemands	43	22.52 (3.61)	abs	2	1	1 + 0
	4	participants recrutés en ligne	584	32.14 (12.04)	abs	2	1	1 + 0
Samland et al. (2016)	1	enfants de 5 ans et adultes	150	65.63 mois (3.64) ; 35.37 (10.03)	abs	2	2	1
	2	enfants de 5 ans et adultes	97	64.23 mois (3.02) ; 25.06 (4.51)	abs	2	2	1 + 0
Sarin et al. (2017)	3	abs	42	28.76 (9.25)	14/28	1	2	1
Schleifer et al. (1983)	2	enfants de 5, 7, 9 et 11 ans d'une école de Montréal (francophones)	32	abs	16/16	1 + 3	2	1

Note. Type de question (1 = ouverte ; 2 = centrée sur l'agent) ; Réplication de l'effet de la valence morale (1 = Effet répliqué ; 0 = Effet non répliqué)

Annexe F

Figure F1

Distributions des scores d'évaluation en fonction des motivations de l'agent dans l'Étude 5.



Note. Echelle de 0 à 10 (où 0 = Très négatif ; 5 = neutre ; 10 = Très positif)

Résultat du test de Levene : $F(1, 118) = 2.42, p = .123$ (homogénéité des variances).