Les processus de prise de décision morale chez les personnes borderlines :

Comment l'empathie et la perspective du participant prédisent leurs choix moraux ?

Thèse présentée le 13 Novembre 2023 par

Julian NASELLO

En vue de l'obtention du grade académique de

Doctorat en Sciences Psychologiques et de l'Education.

Composition du Jury :

Pr. Jean-Marc Triffaux (Promoteur), **Pr. Adélaïde Blavier** (Co-promotrice),

Pr. Michel Hansenne (Président du Jury), **Pr. Benoît Dardenne** (Secrétaire du Jury),

Pr. Olivier Luminet (Membre externe du Jury) et **Pr. Serge Tisseron** (Membre externe du Jury).

Résumé

Est-il préférable de sauver cinq personnes et d'en sacrifier une seule, ou l'inverse ? Si cette question vous semble familière c'est parce qu'elle fait référence à un dilemme bien connu en philosophie, le problème du trolley. Ce dernier oppose deux courants : l'utilitarisme et le déontologisme. Le premier principe est centré sur les conséquences d'une action et sur la maximisation du bien-être ou la minimisation du mal-être ; le second est centré sur l'action en elle-même : si elle viole ou non les normes morales.

Dans un cadre expérimental, nous pouvons ainsi étudier les prises de décision morales des individus en utilisant des dilemmes de type trolley à leurs choix : utilitaires (i.e., choisir de sacrifier la personne isolée pour sauver les cinq) ou déontologiques (i.e., choisir de sacrifier les cinq pour sauver la personne isolée). Un modèle fait actuellement référence pour expliquer les jugements et choix moraux, le *dual-process model* (DPM, Greene et al., 2001 ; 2004). Ce dernier illustre que les situations morales qui élicitent principalement des réponses émotionnelles vont produire davantage de choix déontologiques. Inversement, les situations morales qui engagent peu de réponses émotionnelles vont permettre aux processus cognitifs d'intervenir davantage et de produire alors plus de choix utilitaires.

Sur cette base, deux facteurs apparaissent primordiaux : l'empathie (présentant une facette affective et cognitive) et les traits borderline. En effet, il est connu que de faibles niveaux d'empathie affective prédisent davantage de choix utilitaires. Les traits borderline étant liés significativement à des niveaux d'empathie affective élevés, nous avons tenté de déterminer s'ils prédisaient moins de choix utilitaires.

Dès lors, nous avons réalisé trois études pour tester ces hypothèses, ainsi qu'une revue systématique et méta-analyse de la littérature. La première étude visait à étudier les liens entre l'empathie et des dilemmes de type trolley mettant en scène des situations plus écologiques, tout en invitant le participant à incarner différents protagonistes. La deuxième étude visait à comparer des dilemmes de type trolley classiques à des dilemmes de la vie de tous les jours pour étudier comment l'empathie et les traits borderline prédisaient les choix moraux, tout en invitant le participant à incarner différents protagonistes. La troisième étude visait à répliquer la seconde sur une population clinique de patientes borderline. Enfin, nous avons conduit une méta-analyse et revue systématique de la littérature pour clarifier les liens entre empathie et jugements, choix et inclinaison morale dans les dilemmes de type trolley.

Ce travail a permis de dessiner les limites inhérentes au DPM tout en illustrant que le rôle attribué à l'empathie (voire également aux traits de personnalité de manière générale) dans les jugements, les choix et la préférence morale apparait surévalué. Enfin, nous avons pu constater que le type de dilemme et la perspective du participant dans ces dilemmes avaient un impact majeur sur les prises de décision des individus.

Abstract

Would you save five individuals and sacrifice one, or vice versa? If this question seems familiar to you, it is because it references a well-known philosophical dilemma, the trolley problem. This dilemma pits two philosophical schools of thought against each other: utilitarianism and deontology. The former principle focuses on the consequences of an action and the maximization of well-being or the minimization of suffering, while the latter centers on the action itself: whether it violates moral norms or not.

In an experimental context, we thus examine individuals' moral decision-making by using trolley-type dilemmas in which they have to choose between utilitarian options (i.e., choosing to sacrifice the isolated individual to save the five) or deontological options (i.e., choosing to sacrifice the five to save the isolated individual). A model currently referenced for explaining moral judgments and choices is the dual-process model (DPM, Greene et al., 2001; 2004). This model illustrates that moral situations that primarily elicit emotional responses will produce more deontological choices. Conversely, moral situations that engage fewer emotional responses will allow cognitive processes to intervene more and thus produce more utilitarian choices.

Based on this, two factors appear paramount: empathy (comprising both affective and cognitive facets) and borderline traits. Indeed, it is known that lower levels of affective empathy predict more utilitarian choices. As borderline traits are significantly associated with higher levels of affective empathy, we attempted to determine if they predicted fewer utilitarian choices.

Consequently, we conducted three studies to test these hypotheses and conducted a systematic review and meta-analysis of the literature. The first study aimed to examine the links between empathy and trolley-type dilemmas involving more ecologically valid scenarios, all while inviting the participant to embody different protagonists. The second study aimed to compare classic trolley-type dilemmas with everyday life dilemmas to study how empathy and borderline traits predicted moral choices, all while inviting the participant to embody different protagonists. The third study aimed to replicate the second study with a clinical population of female borderline patients. Finally, we conducted a meta-analysis and systematic review of the literature to clarify the links between empathy and judgments, choices, and moral inclinations in trolley-type dilemmas.

This work has helped delineate the inherent limits of the DPM while illustrating that the role ascribed to empathy (and even personality traits in general) in judgments, choices, and moral preferences appears overestimated. Finally, we observed that the type of dilemma and the participant's perspective in these dilemmas had a major impact on individuals' decision-making.

Remerciements

Je souhaiterais adresser de nombreux remerciements aux personnes qui m'ont accompagné au cours de ce projet. L'aboutissement de cette thèse fut une longue aventure périlleuse mais particulièrement enrichissante.

Mes premiers remerciements vont à mon promoteur principal, le Professeur Jean-Marc Triffaux, et à ma co-promotrice, la Professeure Adélaïde Blavier. Votre soutien et vos encouragements constants m'ont été d'une aide incommensurable dans le cadre de ce travail, mais également dans ma carrière professionnelle. J'imagine à quel point il est complexe d'encadrer un jeune chercheur et je vous remercie du fond du cœur pour vos conseils, votre disponibilité, votre expertise et votre bienveillance.

Je remercie également les Professeurs Michel Hansenne et Benoit Dardenne, deux accompagnants de très longue date étant donné qu'ils m'encadraient déjà pour mon mémoire de fin de master. Merci également du fond du cœur pour votre patience et pour m'avoir fait bénéficier de votre expertise de chercheurs aguerris. J'ai parfois passé des quarts d'heure difficiles à vos côtés, mais toujours à bon escient. Vous avez été (et êtes toujours) deux professeurs talentueux qui m'ont été chers au cours de mon parcours académique. Merci pour tout.

Je remercie le Professeur Etienne Quertemont qui m'a également accompagné au début de mon projet de thèse. Il a été un acteur important dans le cadre de ce projet. Merci énormément pour votre disponibilité et vos précieux conseils.

Je souhaite également remercier le Professeur Francis Pérée pour son immense talent en statistiques et pour ses qualités didactiques. Grâce à vous, j'ai pu progresser énormément dans le domaine des statistiques.

Je remercie enfin les deux membres externes de mon jury de thèse, les Professeurs Serge Tisseron et Olivier Luminet. Merci pour le temps et l'analyse que vous allez consacrer à ce travail, j'espère de tout cœur qu'il correspondra à vos attentes.

Je souhaite remercier chaleureusement tous les participants et participantes de mes études, sans qui tout cela ne serait évidemment possible. Je remercie particulièrement les patientes qui ont acceptées de participer à ma dernière étude, notamment pour le fait

qu’elles m’aient livré des aspects particulièrement difficiles de leur parcours. J’espère que les résultats de ce travail permettront de vous aider.

Je remercie infiniment mes amis et ma famille pour leur aide permanente et leurs participations constantes à mes recherches. Michael Antoine et Samuel Carrio Montes, Céline Michel, mais aussi les Docteurs Youssef Lakhal et Maxime Lognoul ; vos avis, vos conseils et votre soutien font de vous les amis qui me sont le plus chers. Merci également à ma sœur, Cathy Nasello, mes cousins et cousines, Fabio et Linda Bellavia, Anthony et Sarah Bucchieri, ma tante adorée, Franca Nasello, pour votre présence, votre soutien et votre enthousiasme à l’égard de mes travaux. Je vous aime du fond du cœur.

Enfin, le meilleur pour la fin. Je remercie les deux êtres qui me sont le plus chers au monde, les plus belles personnes qui puissent exister. Il s’agit de ma future épouse, Amandine Rosu, et de ma grande fille, Naomi Nasello. Vos sourires et votre amour m’ont permis de traverser les adversités de ce travail, mais aussi de la vie en général, pour devenir un homme meilleur. J’espère vous rendre ne fut-ce qu’une infime partie du bonheur que vous m’apportez au quotidien. Je suis extrêmement fier de vous et je vous aime d’un amour infini.

Table des matières

Section 1 : Background théorique

Chapitre 1 : introduction théorique

Background

Peut-on sacrifier un individu ou un nombre restreint de personnes pour en sauver un nombre plus conséquent ? Cette question a longtemps tourmenté les philosophes de la morale. En effet, pour ne citer que les plus connus, Bentham (1789), Mill (1879) et Sidgwick (1874) ont longtemps défendu l'idée que les actions morales doivent être prescrites selon le *Greatest Happiness Principle*. Ce principe, dit de l'utilitarisme, explique que les actions sont moralement acceptables lorsqu'elles produisent le plus grand bonheur chez le plus grand nombre. De ce fait, l'action doit avoir pour but de réduire la douleur et d'augmenter le plaisir du plus grand nombre. La position utilitaire, visant donc à maximiser *l'utilité*[1], a longtemps été en opposition au déontologisme. Cette position, notamment portée par Kant (1785), défend l'idée qu'il n'est pas possible de quantifier une quantité de plaisir ou de douleur et réfute le fait que les actions morales sont déterminées par la raison. Avec la position utilitaire, on centre la justesse de l'action morale au niveau des conséquences qu'elle produit (cette position est d'ailleurs appelée « conséquentialiste »), tandis qu'avec la position déontologique, c'est l'action elle-même qui est évaluée. En effet, du point de vue déontologique, si l'action morale viole des principes fondamentaux (e.g., tuer intentionnellement un individu), alors cette action doit être proscrite. Du point de vue utilitaire, si cette action produit un effet dont les conséquences vont bénéficier au plus grand nombre, alors l'action doit être réalisée.

Bien plus tard, Foot (1967) et, quelques années après, Thomson (1976), ont créé deux problèmes que l'on a nommés « Dilemmes du Trolley », illustrant parfaitement comment s'articulent les positions utilitaires et déontologiques. Le scénario imaginé par Philippa Foot (1967) se déroule comme ceci : imaginez un train filant à vive allure sur une voie ferrée. Sur la voie principale se trouvent cinq travailleurs ferroviaires qui n'ont pas conscience du danger imminent. En effet, si rien n'est fait, le train va les percuter et ils mourront inévitablement. En tant que conducteur, vous avez la possibilité de dévier le train sur une autre voie où il n'y a qu'un seul travailleur. Si vous prenez la décision de dévier le

[1] L'utilité, comme toutes les notions élémentaires, est assez complexe à définir. Dans sa forme la plus rudimentaire, certains parlent d'une évaluation entre les coûts et les bénéfices d'une chose, en faveur des bénéfices (Powell, 2022), mais elle est également définie comme un « bien, intérêt (de quelqu'un) ; avantage, profit (tiré par quelqu'un de quelque chose) » (CNRTL).

train, c'est le travailleur isolé qui sera dès lors percuté et qui mourra. En sachant qu'il n'y a aucun moyen de freiner ou d'avertir ces travailleurs ferroviaires, que faire ? La version modifiée de Judith Thomson (que l'on nomme le « Footbridge Trolley Problem ») vous transpose sur un pont piéton (*Footbridge*). En regardant en contrebas de ce pont, vous voyez arriver un train à vive allure qui, une fois encore, va percuter inéluctablement cinq ouvriers sur la voie. Paniqué, vous constatez qu'à côté de vous se trouve une personne obèse. En la poussant au trépas, cette personne arrêtera le train dans sa course effrénée ; ce qui sauvera les cinq ouvriers. Dans ce scénario, le choix est identique à la version de Foot (i.e., sacrifier une personne pour en sauver cinq, ou l'inverse), mais l'action est différente. Effectivement, dans la version de Thomson, il faut violer un principe moral (i.e., tuer intentionnellement un individu) pour en sauver un plus grand nombre.

Ces concepts philosophiques ont été largement popularisés à la suite de publications majeures de Greene et collaborateurs (2001 ; 2004 ; Greene, 2007). Ces auteurs ont développé un modèle, le *Dual-Process Model* (DPM), montrant que les processus émotionnels et cognitifs interviennent de manière opposée dans les jugements moraux. En effet, lorsque l'on présente la version de Foot (1967) à des participants, Greene et collaborateurs ont constaté que cela implique davantage de processus cognitifs et les structures cérébrales qui leur sont associées, et nettement moins les processus émotionnels. De cela s'ensuit, avec une plus grande propension, un jugement moral de type utilitaire. À l'inverse, quand les participants sont confrontés à la version de Thomson (1976), on constate que les processus émotionnels sont nettement plus sollicités que les processus cognitifs, et de cela s'ensuit, avec une plus grande propension, un jugement moral de type déontologique.

Avant de rentrer dans le vif du sujet, il faut bien distinguer deux concepts fondamentalement distincts dans la psychologie de la morale : le jugement moral et la prise de décision morale. Comme nous l'avons défini dans une publication récente (Nasello et al., 2023a), décider, c'est mettre fin à un processus de délibération séquentiel. L'individu va sélectionner une option parmi plusieurs, mutuellement exclusives. En revanche, un jugement est un processus d'évaluation. Dans cette évaluation, l'individu va caractériser des actions ou des caractéristiques d'une autre personne comme bonnes ou mauvaises dans une situation particulière (Haidt, 2001). De ce fait, le jugement moral va faire partie du processus de délibération séquentiel et peut contribuer significativement à la prise de

décision morale, mais cette association n'est pas systématiquement significative. En effet, on constate que certaines psychopathologies (e.g., la psychopathie) outrepassent les jugements moraux dans leur prise de décision (Cima et al., 2010 ; Pletti et al., 2017). Dès lors, les personnes psychopathes sont tout à fait capables de discerner le bien du mal, mais n'en tiennent pas compte dans leur prise de décision finale. De plus, les caractéristiques inhérentes d'un dilemme peuvent également conduire à ce que l'on bypasse les jugements moraux. Très concrètement, prenons le dilemme du trolley de Foot (1967) : nous savons qu'une majorité d'individus choisissent l'option utilitaire (environ 90 %), c'est-à-dire dévier le train pour sauver les cinq ouvriers et tuer l'individu isolé. Plaçons à présent une personne chère au décideur sur la voie ferrée alternative ; dans ce cas de figure, une majorité d'individus refusera de détourner le train pour sauver les cinq ouvriers. À travers ces exemples, il apparait clairement que les éléments du dilemme moral et/ou les traits de personnalité du décideur vont faire varier fortement l'association entre jugement moral et prise de décision morale.

Les limites des dilemmes de type trolley dits « classiques »

Bauman et collaborateurs (2014) ont émis de vives critiques à l'encontre des dilemmes de type trolley, en plein foisonnement dans la littérature scientifique. En effet, ces auteurs décrivent ces problèmes comme « amusants » pour les participants plutôt que « préoccupants », ce qui diminue le caractère immersif de ces scénarios. De plus, ils sont jugés comme étant particulièrement éloignés de situations « réelles » ou, en tous les cas, ils ne sont pas considérés comme étant proches de la réalité. Ces critiques décisives ont particulièrement mis à mal la généralisation supposée de ces dilemmes à des situations réelles, impossible à tester pour des raisons éthiques évidentes.

Pour répondre à ces points cruciaux, les auteurs ont élaboré différentes approches. La première a été celle de Navarrette et collaborateurs (2012) qui ont implémenté le dilemme du trolley en réalité virtuelle. Avec un casque immersif, les participants se voyaient incarner un conducteur de tram qui allait percuter ce malheureux groupe de cinq ouvriers ferroviaires. Les sujets avaient pour seule alternative le fait de dévier ce train sur une autre voie où, comme vous le savez, se trouvait un travailleur isolé. Dans cette situation virtuelle, les auteurs ont testé deux conditions expérimentales : « omission » et « action ». Dans la condition *omission*, les participants ne devaient rien faire s'ils souhaitaient sacrifier l'individu isolé pour sauver les cinq autres. À l'inverse, dans la condition *action*, les

participants devaient activer un levier pour dévier le train afin de sacrifier le travailleur isolé. Cette manipulation expérimentale a permis de montrer que les pourcentages de décision morale ne variaient pas par rapport aux dilemmes de type « papier-crayon » et qu'il n'y avait pas de différence entre les conditions *omission* et *action* (c'est-à-dire qu'on retrouve environ 90 % de choix utilitaires dans les deux conditions).

La seconde approche a été de développer des dilemmes de type trolley de la vie de tous les jours. C'est dans cette approche que nous avons inscrit nos travaux. De nombreuses équipes de chercheurs se sont donc attelées à développer tout un ensemble de situations de la vie quotidienne impliquant des dilemmes de type trolley (e.g., Lotto et al., 2014 ; Vyas et al., 2017 ; Pletti et al., 2017 ; Takamatsu, 2019 ; Nasello et al., 2021a ; Nasello et al., 2021a ; Nasello et al., 2023a). Ces derniers sont préoccupants pour les participants (i.e., ils se décrivent comme étant tourmentés par la situation du scénario), ils ne sont pas amusés et décrivent le scénario comme immersif. Dès lors, ces nouveaux attributs en font des outils plus généralisables par rapport aux dilemmes classiques (nous illustrerons ces scénarios dans les prochaines sections).

Quels sont les mécanismes qui prédisent les choix moraux dans les dilemmes de type trolley ?

Comme mentionné, le modèle qui a reçu le plus de support empirique est celui de Greene et al. (2001 ; 2004 ; Greene, 2007). Le DPM met en contraste des processus émotionnels et cognitifs, où lorsque les processus émotionnels apparaissent trop intenses, ils empêchent la survenue des processus cognitifs. Initialement, il a été développé comme modèle explicatif des jugements moraux puis a été étendu aux prises de décisions morales. Les auteurs ont d'ailleurs décrit le dilemme de Foot (1967) comme *impersonnel* car il implique davantage de processus cognitifs et moins de processus émotionnels (dans ce type de dilemme, les processus cognitifs prennent le pas sur les processus émotionnels) ; et celui de Thomson comme *personnel* car il implique davantage de processus émotionnels et moins de processus cognitifs. Des auteurs comme Choe and Min (2011) ont montré que les émotions jouaient bien un rôle significatif dans les jugements moraux (i.e., la colère, la tristesse, le dégoût, l'empathie [classifiée comme une émotion], l'anxiété et surtout le dégoût).

Assez naturellement, d'autres auteurs se sont tournés vers des populations psychopathologiques pour tester la validité du DPM. L'une des plus populaires a été la psychopathie[2], notamment pour l'émoussement affectif qui caractérise cette psychopathologie. En effet, de nombreux auteurs (Koenigs et al., 2012 ; Gao and Tang, 2013 ; Patil, 2015 ; Balash and Falkenbach, 2018) ont montré que les personnes psychopathes présentaient une nette propension aux choix utilitaires, en comparaison avec des individus « contrôles » ; ce qui a renforcé le pouvoir prédictif du DPM.

Toutefois, malgré ces résultats apportant de la consistance au DPM, de plus en plus d'auteurs ont commencé à questionner la validité du modèle (Horne & Powell, 2016 ; Smillie et al., 2020 ; Decety, 2021 ; Bègue et al., 2015 ; Oudman et al., 2021). En effet, certains auteurs ont recensé une taille d'effet particulièrement faible du rôle prétendument prépondérant des processus émotionnels et cognitifs.

Par ailleurs, nous aborderons deux positions qui s'affrontent dans l'une de nos publications (Nasello & Triffaux, 2023b), d'un côté des auteurs qui vont considérer que les émotions sont responsables du temple de la moralité (Haidt, 2003) et d'autres vont soutenir l'idée que nous manquons de preuves pour conclure qu'elles jouent un rôle majeur dans la morale (e.g., Horne & Powell, 2016 ; May, 2018). Positions dichotomiques que l'on retrouve également avec l'empathie (Decety, 2021 ; Decety & Cowell, 2015 ; de Waal et al., 2006).

Comme mentionné, certains auteurs se sont tournés vers la psychopathologie pour tester la validité du DPM et pour mieux comprendre les mécanismes sous-jacents aux décisions et jugements moraux. D'autres ont investigué l'influence de certaines dispositions comme l'empathie ou l'alexithymie (e.g., Patil & Silani, 2014 ; Zhang et al., 2020).

L'empathie : définition et modélisation

L'empathie est un phénomène qui a gagné en intérêt auprès de la communauté scientifique depuis ces 20 dernières années. Bien qu'elle soit fondamentale, elle est relativement complexe à définir. Classiquement, l'empathie est subdivisée en deux composantes : une composante affective qui correspond à la capacité à ressentir les

[2] D'autres populations cliniques ont également été testées, notamment les troubles du spectre autistique (Gleichgerrcht et al., 2013 ; Patil et al., 2016 ; Li and Liu, 2017), les troubles anxieux (Patil et al., 2021) et le syndrome Korsakoff (Oudman et al., 2021).

émotions des autres (Bryant, 1982) et une composante cognitive correspondant à la capacité de comprendre les émotions des autres (Hogan, 1969), capacité qui va être étendue à tous les états mentaux confondus par certains auteurs (Davis, 1980 ; Davis, 1983). A l'heure actuelle, il y a un débat toujours ouvert par rapport à la dimension cognitive de l'empathie : s'agit-il de comprendre uniquement les aspects émotionnels des autres ou plutôt tous les états mentaux (i.e., ses pensées d'autrui, sa vie mentale et ses émotions) ? On ajoute aussi à l'empathie une composante motivationnelle décrite comme le fait d'être préoccupé à propos d'autrui ou motivé à se soucier de leur bien-être (Decety, 2016). Selon la définition que l'on en donne, il existe différentes échelles de mesure dont les plus connues sont l'*Interpersonal Reactivity Index* (Davis, 1983), la *Jefferson Scale of Empathy* (Hojat, 2002) et la *Basic Empathy Scale* (Jolliffe & Farrington, 2006).

Dès lors, pour quelles raisons l'empathie reste-t-elle une caractéristique difficile à cerner ? D'une part, une méta-analyse assez récente a montré, qu'au sein de la population générale, les individus adhéraient à des thèmes spécifiques de l'empathie (des définitions spécifiques) très faiblement corrélés entre eux (Eklund & Meranius, 2021). En d'autres mots, la plupart des individus vont attribuer à l'empathie des thèmes particuliers, en excluant ou en ne prenant pas en considération ses autres aspects. Ceci signifie que, dans l'ensemble, les gens ont une vision très partielle et limitée de l'empathie. Les auteurs de cette méta-analyse ont recensé quatre thèmes principaux liées à l'empathie : comprendre (i.e., la capacité d'avoir une certaine appréhension de la vie mentale d'une autre personne), ressentir (i.e., avoir une réponse émotionnelle appropriée à la situation vécue par une autre personne), partager (i.e., faire l'expérience d'états similaires à ceux d'une autre personne) et enfin, différencier soi et autrui (i.e., reconnaitre qu'il y a une différence entre soi et les autres ; entre une source et soi).

D'autre part, une autre méta-analyse a recensé 43 définitions de l'empathie dans la littérature, montrant la très large variété des points de vue d'experts autour de ce phénomène (Cuff et al., 2016). Les auteurs ont alors proposé la définition synthétique suivante : « *l'empathie est une réponse émotionnelle (affective), dépendante de capacités « traits » et d'influences « états ». Les processus empathiques sont déclenchés automatiquement mais sont également façonnés par des processus dits « top-down ». L'émotion qui résulte* (chez l'observateur)*, est similaire à la perception qu'il s'en fait (directe ou imaginée) et à la compréhension du stimulus émotionnel (empathie cognitive).*

Enfin, il y a une reconnaissance que la source de l'émotion n'émane pas de soi » (Cuff et al., 2016, p. 150).

Afin d'illustrer la complexité du phénomène, nous avons modélisé l'articulation des différentes facettes de l'empathie dans une publication (Nasello & Triffaux, 2023a). Décrite comme un processus, on constate qu'il existe différentes voies qui vont produire des réponses spécifiques chez l'empathiseur (i.e., l'observateur) selon que certaines facettes soient plus ou moins proéminentes chez ce dernier.

En résumé, ce modèle nous montre que la facette affective de l'empathie correspond à ce que l'on nomme « la contagion émotionnelle ». Cette caractéristique étant automatique, elle peut entrainer tout un processus de contamination émotionnelle chez l'empathiseur si l'émotion n'est pas régulée de façon adéquate (i.e., lorsque de mauvaises stratégies de régulation émotionnelle sont employées) ; ce qui pourrait conduire à une submersion ou surcharge émotionnelle chez l'empathiseur. Dans cette situation, ce dernier présentera davantage de chance de produire des comportements inappropriés (e.g., dramatiser la situation de l'empathisé [la cible]). Inversement, cumuler l'emploi d'une mauvaise stratégie de régulation émotionnelle (e.g., la suppression, le déni ou la rumination) et une faible motivation à se soucier de l'autre va entrainer des attitudes froides ou de l'indifférence chez l'empathiseur envers l'empathisé.

Dès lors, ce modèle positionne la régulation émotionnelle au centre du processus empathique car les stratégies de régulation qui seront employées par l'empathiseur vont catalyser des processus différents.

L'empathie cognitive a, dans ce modèle, une interaction directe avec la motivation à aider (à se soucier de l'autre). C'est lorsque l'empathiseur va réguler ses états émotionnels de façon appropriée et présenter une motivation à aider adéquate qu'il va pouvoir se mettre à la place de l'empathisé pour mieux comprendre l'état de celui-ci (empathie cognitive). Une trop forte propension à se mettre à la place d'autrui peut augmenter l'identification à autrui, ce qui peut réenclencher un processus de contamination émotionnelle (i.e., si l'empathiseur s'identifie à l'empathisé, il y aura une motivation à aider plus importante qui, si trop élevée, peut entrainer une submersion émotionnelle). Trop faiblement développée, l'empathiseur qui manque d'habilités à comprendre les émotions ou la situation de

l'empathisé, tout en étant adéquatement motivé(e) à se soucier de l'autre, collectera de façon probablement insuffisante des informations critiques chez l'empathisé.

Ainsi, comme nous le présentons à travers ce modèle, il est nécessaire de développer des stratégies de régulation émotionnelle adéquates et adaptatives (e.g., l'acceptation, la centration positive, la centration sur l'action, la réévaluation positive et la mise en perspective), tout en présentant un souci de l'autre et une prise de perspective adaptée.

L'article ci-après illustre ce processus en détails.

Article : A stenography of empathy: Toward a consensual model of the empathic process

L'Encéphale xxx (xxxx) xxx–xxx

Disponible en ligne sur
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Literature review

A stenography of empathy: Toward a consensual model of the empathic process

Une sténographie de l'empathie : vers un modèle consensuel du processus empathique

J.A. Nasello [a,b,*], J.-M. Triffaux [a,c]

[a] *Psychiatric Day Hospital "La Clé", boulevard de la Constitution 153, 4020 Liège, Belgium*
[b] *Department of Clinical Psychology, University of Liège, place des Orateurs 1, 4000 Liège, Belgium*
[c] *Department of Psychiatry, Medicine, University of Liège, CHU of Liège, avenue de L'Hôptial 1, 4000 Liège, Belgium*

ARTICLE INFO

Article history:
Received 29 August 2022
Accepted 9 December 2022
Available online xxx

Keywords:
Empathy
Emotional contagion
Emotion regulation
Psychopathology
Education

ABSTRACT

Empathy has gained popularity in the general population and the scientific world during the past decade. Recently, several researchers found a significant decrease in empathy scores of healthcare students (notably medical students) and recommend promoting empathy skills in several fields of education. The current paper presents a new model of the empathic process: a stenography of empathy compelling scientific data and contemporary conceptions. Indeed, we combined all pioneer researchers' conceptions of empathy (Davis, Decety, Batson, Preston & de Waal) into an integrative model. This model is centered on the empathizer (i.e., a person observing a target experiencing emotions) and displays how all empathy components are articulated, explaining the individuals' general functioning and how the process might become dysfunctional. We illustrated applications of the model with three clinical examples (i.e., burnout, psychopathy, and borderline personality disorders) to display how empathy is related to psychopathological symptoms. We believe this new dynamic and sequential model would be helpful in explaining how empathy works, which is of great interest to healthcare students, clinicians, researchers, and academics.

Mots clés :
Empathie
Contagion émotionnelle
Régulation émotionnelle
Psychopathologie
Formation académique

RÉSUMÉ

Depuis ces deux dernières décennies, on observe un intérêt croissant au sujet de l'empathie, tant au sein de la population générale que dans le monde scientifique. Effectivement, plus de la moitié des productions scientifiques sur l'empathie ont été réalisées depuis 2012. Récemment, de nombreuses études se sont focalisées sur le développement de l'empathie des étudiants (notamment en médecine), de par l'observation d'un déclin significatif de leurs scores au cours de leur parcours académique. Si la plupart de ces travaux prônent la nécessité d'un développement accru des compétences empathiques, certes, de nouveaux supports théoriques méritent d'être construits afin d'en faciliter l'apprentissage théorico-pratique. Le présent article propose une nouvelle modélisation du processus empathique. Il s'agit d'une sténographie de l'empathie car il se rapporte à une simplification d'un phénomène complexe aux frontières floues, voire parfois ineffables. Ce modèle intégratif combine les données scientifiques contemporaines et les différentes approches théoriques d'auteurs pionniers du domaine comme M. Davis, J. Decety, C. Batson, F. de Waal et S. Preston. Centré sur « l'empathiseur » (c.-à-d., la personne qui observe une cible éprouvant une émotion), le modèle décrit comment les composantes de l'empathie s'articulent entre elles. Il permet d'expliquer le fonctionnement général des individus mais aussi comment le processus peut devenir

* Corresponding author at: Psychiatric Day Hospital "La Clé", boulevard de la Constitution 153, 4020 Liège, Belgium.
E-mail address: julian.nasello@gmail.com (J.A. Nasello).

https://doi.org/10.1016/j.encep.2022.12.002
0013-7006/© 2023 L'Encéphale, Paris.

dysfonctionnel. Dans ce présent article, nous avons également illustré à travers trois exemples cliniques (c.-à-d., le *burnout*, la psychopathie et le trouble de personnalité *borderline*), comment l'empathie peut être associée à des symptômes psychopathologiques comme le cynisme ou l'épuisement émotionnel, ou encore, les comportements antisociaux et l'émoussement affectif. Nous pensons que ce modèle dynamique et séquentiel offrira un nouveau background théorique clair et didactique pour expliquer et enseigner le fonctionnement de l'empathie (et ses éventuelles dysfonctions), ce qui reste d'un grand intérêt pour les futurs professionnels de la santé, les cliniciens, les chercheurs et les enseignants.

1. Introduction

People have a growing interest in empathy (cf. Google Trends©), and half of the articles on empathy in psychology have been published since 2012, showing that this interest is recent [1]. Currently, students have been in the spotlight because several authors propose to improve their empathy skills in almost all disciplines (medicine [2,3], nursing [4,5], pharmacy [6], health sciences [7], engineering [8], and project management [9]). These studies are particularly right to propose improving empathy abilities given that numerous studies have shown a significant decrease throughout the academic career (especially in medicine [10–12]). However, what exactly do we need to improve?

1.1. What is empathy?

Cuff et al. identified 43 definitions of empathy and synthesized them into one: "empathy is an emotional response (affective), dependent upon the interaction between trait capacities and state influences. Empathic processes are automatically elicited but are also shaped by top-down control processes. The resulting emotion is similar to one's perception (directly experienced or imagined) and understanding (cognitive empathy) of the stimulus emotion, with a recognition that the source of the emotion is not one's own." [13, p. 150].

More recently, Eklund and Meranius found four main themes in the empathy literature, drawing a consensus concerning the empathizer (i.e., the observer) [1]. The first theme is *understanding* and refers to knowing something about the mental life of another person. The second theme is *feeling* which corresponds to an appropriate affective response to another person's situation. The third theme is *sharing* (i.e., experiencing similar states to those the other person is experiencing). Finally, the fourth theme is *self-other differentiation*, which refers to recognizing a differentiation between the other person and oneself.

These clarifications are necessary because they help move forward through a consistent theoretical empathy framework. In line with these improvements, the current paper proposes articulating several conceptions of empathy into a sequential model of the empathic process. Several authors recommended promoting empathy skills (notably for health care professionals), but, as mentioned by Triffaux et al., we need to propose tailored interventions on empathy [14]. For instance, as they mentioned, if we use Davis' theoretical framework [15], would it be necessary to promote personal distress and fantasy? Therefore, before promoting empathy skills, it is important to rely on theoretical models offering a better understanding of how empathy works (e.g., to target specific empathy domains).

In the present paper, we propose a new model that is expected to have developmental, educational, and clinical implications. Indeed, we believe this modelization will offer a clearer view of the empathic process to develop specific empathic abilities (for instance, during the academic curriculum). In addition, clinicians and researchers will be able to identify components of the empathic process that might be related to psychopathology symptoms (e.g., burnout symptoms, psychopathy, or borderline personality disorders, *see* Section 4).

Hence, the model will be a stenography of empathy because it will transcribe a complex phenomenon in a shorthanded way: some ineffable features and connected components form the empathy phenomenon, making all theoretical approaches a simplification of reality.

2. A new model of empathy

The model proposed in the current paper (see Fig. 1) will describe an interaction between an empathizer (i.e., an observer interacting with a target) and an empathee (i.e., the target experiencing an emotional state). The most important characteristic of this interaction is that the emotion felt by the empathee irradiated the empathizer. In other words, no matter what the empathizer does, the emotion felt by the empathee will activate some empathizer's neural patterns; de Waal & Preston called this step *emotion transfer* [16]. This activation has been demonstrated by Preston and de Waal in their Perception-Action model of empathy [17]. They showed that observing or imagining another person's feelings automatically activates some state patterns in the observer. This is an automatic activation because it is elicited unintentionally, requires minimal cognitive resources, cannot be stopped voluntarily, and occurs without conscious awareness [18]. Other authors showed that the observer's brain circuits partially overlap the target's brain circuits when s/he is experiencing a positive or negative emotion [19,20].

Numerous studies showed that observing someone's emotion (or pain) activates similar neural patterns in the observer [21–24]. However, this process is described as selective and influenced by similarity, familiarity, membership, or cooperation [18,25–27], showing that this process is flexible and context-dependent, as suggested by Yamamoto [28]. Parallelly to this automatic activation, the empathizer makes a rapid and brief assessment (RBA) of the empathee (see the RBA component in Fig. 1): s/he identifies the emotional state (e.g., its valence and arousal) and the state of need of the target (i.e., the empathee). For instance, Singer showed that people use representations of their emotional responses to understand others' feelings [29]. This is also supported by the Perception-Action Model and mirror-neuron theories [17]: "*Individuals understand and have a sense of others' emotions because the nervous system evolved to map others' state onto their own individual representations for experiencing those states. As such, when an observer attends to another's state, he or she spontaneously accesses about the other (the 'target'), their feelings, the situation and other related concepts through a distributed associative process*" [16, p. 499].

After this stage, the empathizer enters into a contagion path: the empathee's emotion arouses similar feelings in the empathizer (i.e., a vicarious feeling). Then, this vicarious feeling can evolve through

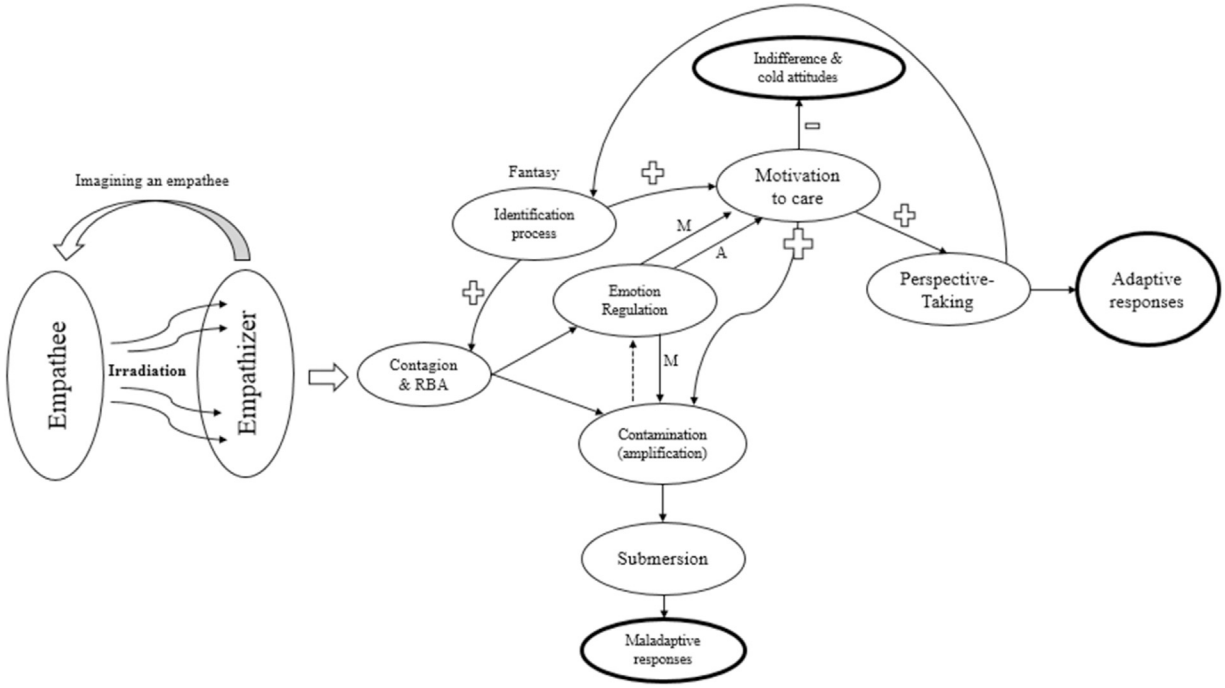


Fig. 1. The modelization of the empathic process. RBA: rapid and brief assessment of the empathee; M: maladaptive emotion regulation; A: adaptive emotion regulation.

contamination (see Fig. 1): an amplification process starts (i.e., the empathizer's emotion gains in intensity), and if the empathizer does not appropriately regulate this amplification (e.g., if maladaptive coping strategies are used), it leads to *submersion* (i.e., the empathizer's emotion reaches an important level). This evolution will mainly produce maladaptive responses[1] such as psychological distress, dramatizing, or overreacting (unproportioned reaction). Dixon-Gordon and collaborators showed that the growing intensity of emotions (especially negative ones) is linked with greater efforts to regulate these emotions [30]. This finding is illustrated by the dotted line between *Emotion Regulation* and *Contamination* in Fig. 1: the empathizer becomes conscious that his/her emotion is amplified and makes more arduous efforts to down-regulate this feeling.

Nevertheless, once the vicarious feeling arises in the empathizer (contagion step), there is another path of evolution. The vicarious feeling is regulated (i.e., transformed into something different, in terms of intensity or emotional state) in a maladaptive or adaptive way. This step involves emotion regulation abilities (i.e., strategies attempting to influence the individual's emotional experiences) and encompasses both positive and negative emotions [31]. In that sense, De Vaus and collaborators stressed that emotions are likely to change: they do not last and are likely to be followed by other emotions [32].

Garnefski et al. identified several cognitive emotion regulation strategies that can be divided into adaptive (e.g., putting into perspective, positive refocusing, positive reappraisal, acceptance, planning) and maladaptive strategies (e.g., self-blame, other-blame, rumination, catastrophizing, suppression) [33]. In addition, Gross identified five families of emotion regulation strategies: situation selection, situation modification, attention deployment, cognitive change, and response modulation [34,35]. The first four are considered *antecedent-focused* (i.e., they occurred before the emotional responses), and the latter *response-focused* (i.e., employed once the emotional response has been activated).

This regulation process leads to several options. As displayed (see Fig. 1), maladaptive emotion regulation can lead to an incongruent emotion (e.g., if the empathee's emotion was sadness, the same emotional pattern is initially elicited in the empathizer, but it quickly evolves into another negative feeling, like angriness). Then, maladaptive emotion regulation can lead to (1) the contamination-submersion path or (2) indifference and cold attitudes if coupled with a low motivation to care.

The optimal path is when the feeling is appropriately regulated (e.g., using acceptance or positive reappraisal strategies). Then, coupled with a high degree of motivation to care (corresponding to empathic concern[2]), it will foster perspective-taking and promote more adaptive responses (e.g., active listening, proper identification of empathee's needs, proposing targeted help, or functional altruistic behaviors). In our modelization, perspective-taking is putting yourself in someone's shoes (by keeping your own socks on) to understand what s/he's is experiencing: the empathizer incorporates and transposes him/herself by simulating the empathee's perspective, but s/he clearly distinguishes him/herself from the empathee (as suggested in systematic reviews [1,13]). This process allows the empathizer to understand the empathee's emotions, goals, and needs by making inferences (i.e., some statistical projections). We based our assumption that perspective-taking is especially involved with adaptive responses on the Russian-doll

[1] Important note: the current model identifies adaptive and maladaptive responses. These responses are qualified as adaptive or maladaptive from a middle to long-term perspective, and from the benefits or damages the empathee will keep from this interaction. For instance, dramatizing or overreacting might be perceived by the empathee as an appropriate response (and maybe more), from a short-term perspective. However, if a clinician repeatedly overreacts in response to the distress of one patient, this patient will learn ineffective emotion regulation strategies in the long run. Moreover, the patient might learn that being distressed attracts attention and encourages people to help him/her, which might entertain his/her pathology.

[2] Empathic concern is frequently associated with sympathy [78]. However, some authors reframed the notion of sympathy by mentioning that sympathy includes empathy, but while sympathy includes going along with a party, empathy would be more unbiased [79]. An empathizer understands but is not especially agree with a particular cause.

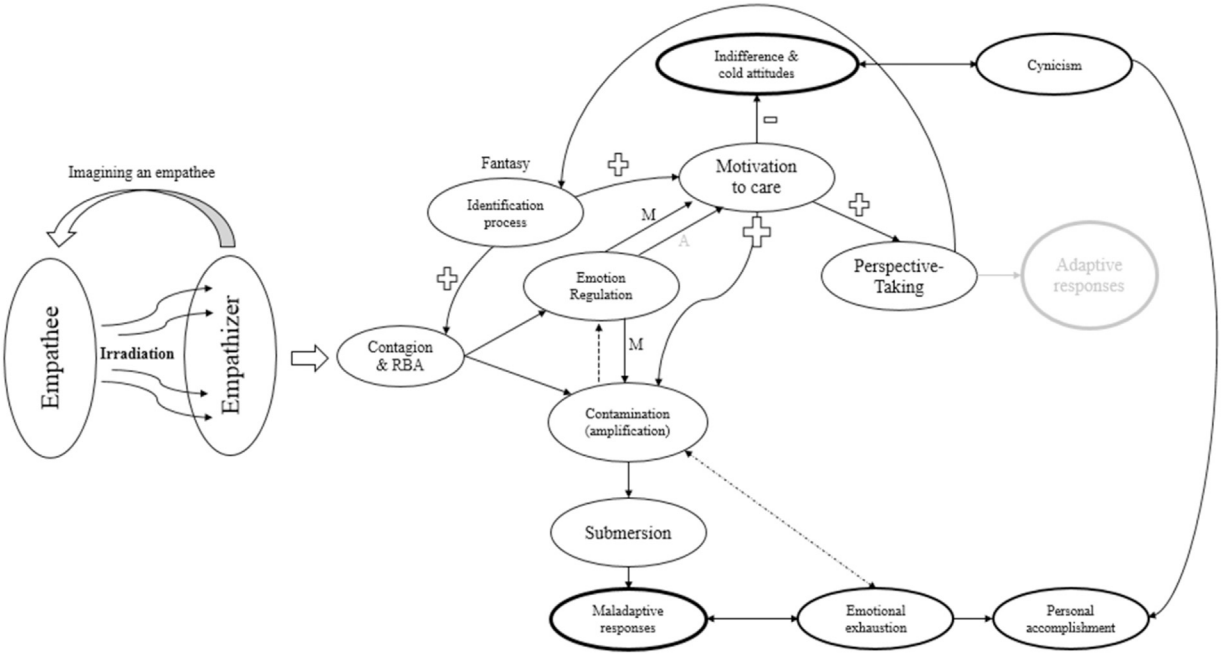


Fig. 2. The empathy model and its associations with burnout domains. RBA: rapid and brief assessment of the empathee; M: maladaptive emotion regulation; A: adaptive emotion regulation.

model of the evolution of empathy [16]. This three-stage model shows that: (1) the perception-action mechanism (motor mimicry) is linked with emotional contagion, (2) empathic concern is linked with consolation, and (3) perspective-taking is related to targeted helping. The downside of perspective-taking is that the empathizer might understand the situation differently when taking the empathee's perspective, triggering the identification process (see Fig. 1) that is fostered by shared features and kinship or ingroup[3] affiliations. Then, when this identification process occurs, the motivation to care increases, leading to the contamination path. Decety and Cowell mentioned that "*explicitly projecting oneself into the patient's situation led to higher levels of personal distress*" [36, p. 8], and we hypothesize that the identification process mediates this effect. Indeed, the personal distress domain of Davis' approach corresponds to the contagion-submersion path (see Fig. 1) [15], and the identification process can exacerbate this path *per se* or when the empathizer takes the perspective of the empathee.

Another important component of our model is the motivation to care. Decety and Cowell mentioned that the motivation to care for others is deeply rooted in our biology, very flexible, and arises early in ontogeny [36, p. 7]. We consider this step the cornerstone component of our model and join Jean Decety's conception that empathic concern refers to the motivation to care for another's welfare [37]. Indeed, depending on the empathizer's degree of motivation, it will foster drastically different responses. For instance, when motivation reaches a too high level (notably when the empathizer strongly identifies or confounds him/herself with the empathee), the empathizer might enter into a contamination process (e.g., the empathizer thinks that there is an emergency to help) and this can lead to submersion and maladaptive responses. Generally, higher motivation to care is influenced by identification (i.e., a cognitive process of emotional participation in which we perceive others as similar to ourselves) and shared features between the empathee and empathizer (e.g., same gender, same past experience, or same age) [38]. As shown by previous findings, *similarity* (i.e., the perceived overlap between one person and another) is recognized as one fundamental factor facilitating empathy [16,17]. In the current paper, we posit that it is especially true for the motivation to care and the contagion process (e.g., if the empathizer perceives that the empathee has the same sportive/political/professional interests, it will potentiate the motivation to care for the empathee and the risk for contagion). However, when the vicarious feeling has elicited another feeling (e.g., sadness has elicited angriness or disdain), if coupled with low motivation to care (see Fig. 1), it can lead to indifference and/or cold attitudes (i.e., a feeling of not being concerned by the empathee's experience, like becoming too insensitive or detached).

The current model places the motivation to care as a cornerstone feature and aligns with Batson's empathy-altruism hypothesis supporting that empathic concern produces altruistic motivation [39–41] (but it extends and nuances specific features of this assumption). According to Batson, *altruism* is a motivational state with the ultimate goal of increasing another's welfare [39, p. 1]; hence, empathy eases behaviors aiming to increase another's welfare. However, there are several ways to consider someone altruistic. As displayed by our model (see Fig. 1), the degree of motivation to care can lead to different outcomes: maladaptive or adaptive responses, but how we interpret these responses (as maladaptive or adaptive[1]) depends on several factors (notably the perspective: from the empathee's, the empathizer's, or the observer's perspective). For instance, an empathee (e.g., a patient) might find it adaptive that the empathizer (e.g., his/her therapist) overreacts or reacts like there is an emergency (and might consider him/her altruistic). However, if this pattern is repeated, it will lead to several issues for the patients (e.g., learning maladaptive emotion regulation strategies) and the therapist (e.g., emotional exhaustion). In other words, the interaction (and interpretation) between the empathic process and altruism is complex and depends on several mechanisms like emotion regulation strategies, motivation to care (empathic concern), and contamination or perspective-taking.

[3] People favor ingroup over outgroup members from a very early age and across different cultures [36,80,81]. This propensity to favor ingroup members is known as the intergroup empathy bias [82].

Lastly, another connex feature linked to the empathic process is the empathizer's ability to create an empathee (e.g., someone s/he anticipates interacting with). In other words, the empathizer mentally simulates an empathee (as proposed by Cuff et al. [13]): s/he imagines a specific situation where the empathee experience a specific emotion (anticipated by the empathizer) and his/her probable reaction.

3. Moderators and external factors influencing empathy

Many moderators and external factors influence our model's components. For instance, the perceived ability to help the empathee successfully is expected to influence the motivation to care [42]. Insofar the empathizer can perceive s/he has insufficient means and capacities to help the empathee (e.g., too much physical distance between the empathizer and empathee, or other professional duties), s/he will be more likely to feel powerless and enter into a contamination process. Inversely, when the empathizer feels s/he has sufficient means and capacities to help the empathee, it will impact motivation and promote perspective-taking and adaptive responses. Here are other examples: (1) other findings showed that emotion perception is context-dependent [28,43], suggesting that context features influence the RBA component (see Fig. 1); (2) the type of communication (e.g., displaying intentional care) is also expected to be a tool that promotes perspective-taking and empathic concern [44]; or (3) the intergroup empathy bias appears to modulate perspective-taking and personal distress [45].

This paper's current purpose is also to invite researchers to test other components that might be added to the model to empower its faculty to explain the empathic process.

4. Practical applications of the empathy model

The model highlights the complexity of the empathic process and stresses that this is a dynamic and sequential process. Moreover, it describes how empathy works and how specific components might become dysfunctional, which is of interest to clinicians, researchers, and professionals aiming to teach empathy.

T. Singer explained that "empathy refers to a complex and multilevel concept incorporating processes of affect sharing, mental state attribution and action control and initiation" [29, p. 858], stressing the importance of benefiting from a model that displays how these parameters are articulated. We also join Preston and de Waal's idea that empathy is a profoundly personalized phenomenon [46]. Indeed, our modelization displays how transitions might occur during this process, leading to a personal configuration for each individual (that will depend on proximal situational features and psychological settings).

The following section will expose three practical applications of our model to explain the associations between empathy and burnout, borderline personality, and psychopathic traits.

4.1. The associations between empathy and burnout

The scientific literature generally describes burnout as a three-domain phenomenon: emotional exhaustion (i.e., feeling emotionally depleted by work, resulting in both psychological and physical symptoms of fatigue); depersonalization or cynicism (i.e., treating others as impersonal objects or diagnoses); and low personal accomplishment (i.e., feeling a lack of intrinsic work-related satisfaction) [47–49]. Based on a systematic review, empathy is related to burnout in healthcare professionals [50]. The authors showed that most studies in their review displayed significant (negative) correlations between empathy and emotional exhaustion, depersonalization, and personal accomplishment. In Fig. 2, we hypothesize that the *indifference and cold attitudes* path (see Fig. 2) is related to cynicism or depersonalization. Based on our model, people display indifference and cold attitudes when they frequently use maladaptive emotion regulation strategies (e.g., suppression) and present low motivation to care about the empathee. For instance, several authors stressed that a gradual emotional (and cognitive) withdrawal leads to progressively displaying detached responses at work and starting to treat others as impersonal objects

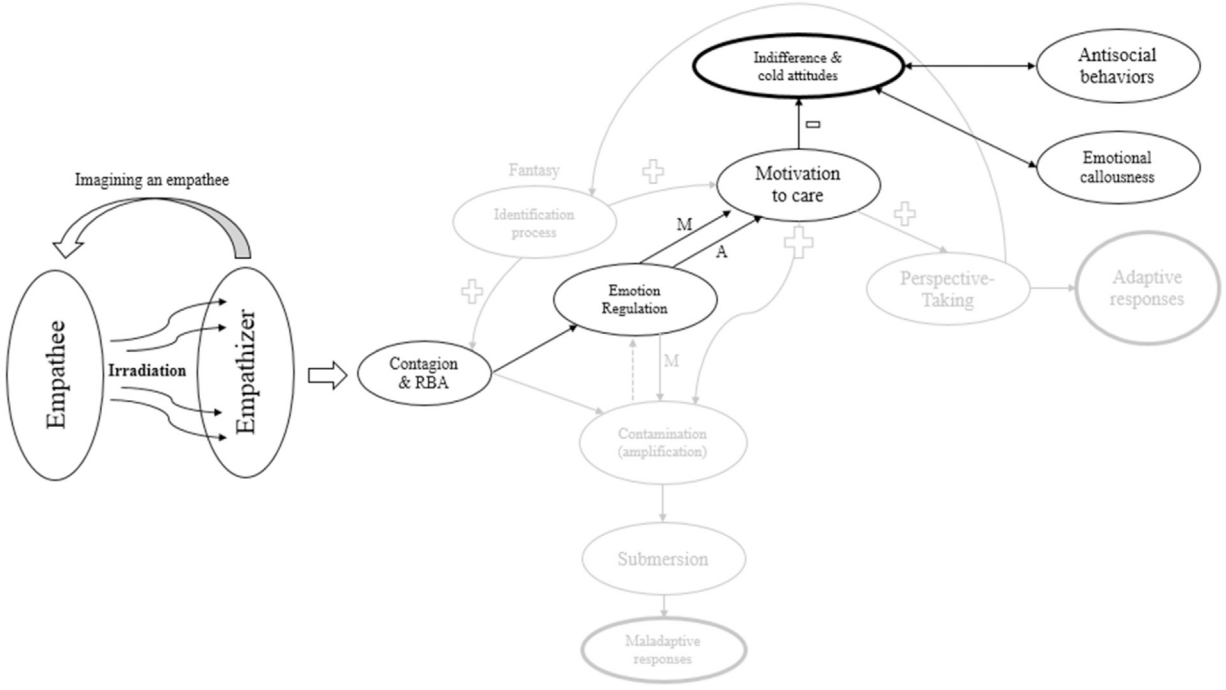


Fig. 3. The empathy model and its associations with psychopathic traits. RBA: rapid and brief assessment of the empathee; M: maladaptive emotion regulation; A: adaptive emotion regulation.

or diagnoses [47,51]. Therefore, we hypothesize that the reoccurrence of a specific path creates a recurrent pattern of interaction that would be, in this case, preferentially related to cynicism (the more indifference and detached affects there will be, the more the individual will be likely to dehumanize others). Then, based on previous findings [52], both cynicism and emotional exhaustion decreased the feeling of personal accomplishment.

Concerning emotional exhaustion, there are several ways to be emotionally exhausted. The first path (see Fig. 2) is when the empathizer does not use efficient emotional regulation strategies (e.g., suppression, self-blame, rumination), and the emotional experience is amplified to the point that this emotion submerses him/her. For instance, Lee and Jang showed that maladaptive emotion regulation strategies (i.e., suppression) significantly increased directly and indirectly emotional exhaustion [53]. Inversely, they showed that adaptive emotion regulation strategies, such as reappraisal, significantly decreased directly and indirectly emotional exhaustion.

The second path is when the empathizer identifies him/herself with the empathee. For instance, Jonsson and Segesten showed that nurses in ambulance services have strong tendencies to identify themselves with the victim, impacting their emotional engagement [38]. In addition, authors found that over-identification with a distressed person was related to higher posttraumatic stress and depression [54,55]. As previously mentioned, the identification process can increase the empathizer's motivation to care and be transformed by an emergency to help the empathee, amplifying the emotional experience and submersion. On the other hand, it is also possible that the motivation to care is appropriate, leading to perspective-taking (see Fig. 2). Then, when the empathizer takes the empathee's perspective, s/he might identify him/herself with the empathee, increasing his/her motivation to care and leading to emotion contamination and submersion. As for cynicism, this repeated interaction pattern would be predominantly related to being emotionally exhausted.

Wilkinson et al. showed that the correlations between empathy and burnout domains vary in intensities and natures (positive or negative), depending on the studies [50]. The current model displayed in Fig. 2 offers explanations of previous findings by showing how people can deploy specific combinations of empathic features that may lead to emotional exhaustion or cynicism.

4.2. The associations between empathy and psychopathic traits

Hare described psychopaths as fearless, callous, and lacking of empathy with impulsive and antisocial behaviors [56]. More precisely, regarding their empathic abilities, studies reported relatively intact cognitive empathy abilities [57,58] but dysfunctions in empathic responding (i.e., personal distress and empathic concern) and deficient responses to transgressions [59,60]. However, a recent meta-analysis [61] showed that psychopathy entertains negative correlations with empathic concern and personal distress (two affective empathy domains) but also with perspective-taking, showing that psychopaths are also impaired in cognitive empathy.

Fig. 3 shows how the empathic process entertains correlations with specific symptoms characterizing psychopathy. Psychopaths are characterized by emotional callousness; hence, they are less triggered by others' emotions [62,63]. As displayed in Fig. 3, they do not enter into a contamination process. Indeed, the vicarious feeling arising from the empathee is quickly "neutralized" (notably because they present fewer activations in response to someone's distress and because they are presumably less sensitive to shared features), and their lack of motivation to care for others (empathic concern) would produce indifference or cold attitudes. Once again, the repetition of these interaction patterns would be predominantly linked with emotional callousness and antisocial behaviors.

4.3. The associations between empathy and borderline personality traits

Borderline personality disorders (BPD) are characterized by pervasive instability of affects, interpersonal relationships, self-image, and marked impulsivity [64]. Regarding BPD patients' empathy, several studies pointed out they present lower cognitive empathy

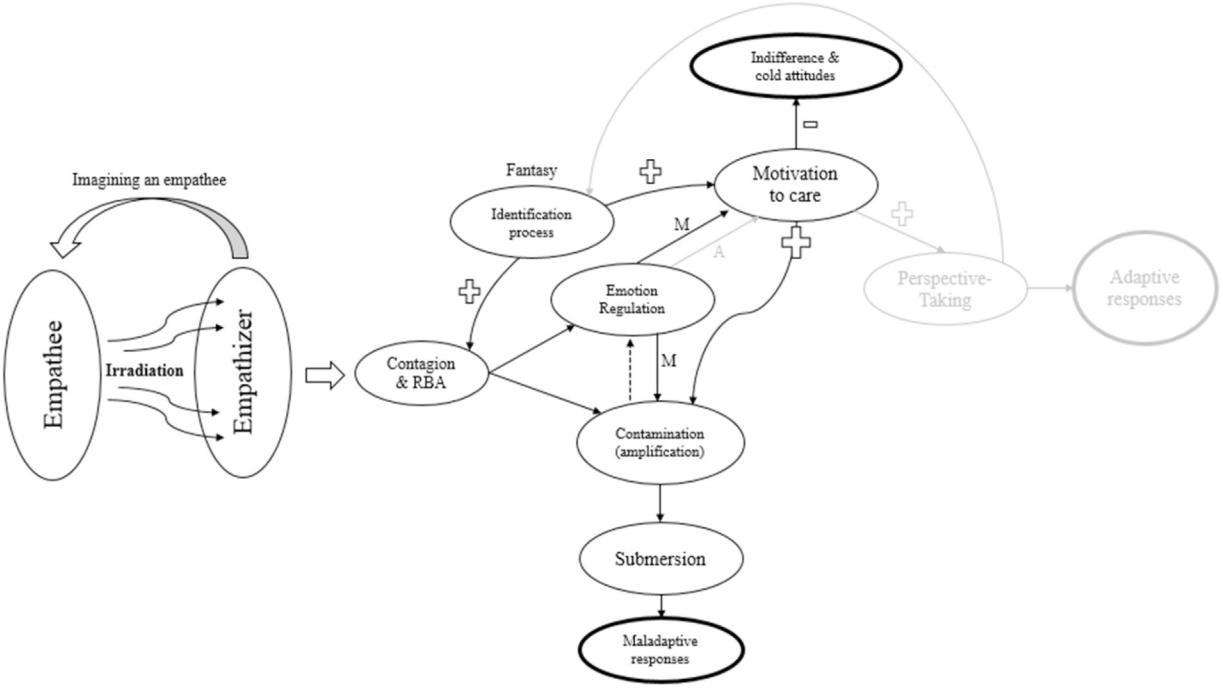


Fig. 4. The empathy model and its associations with borderline personality disorders. RBA: rapid and brief assessment of the empathee; M: maladaptive emotion regulation; A: adaptive emotion regulation.

and higher personal distress (both perspective-taking and personal distress) [65–68] (other studies showed only cognitive empathy [69–71], or personal distress impairments [72,73]). Our model can explain how their empathic process might be related to interpersonal conflicts (see Fig. 4). It is hypothesized that BPD patients predominantly use the contamination to submersion path. Indeed, studies showed that BPD patients display poor emotion management skills and poor emotional understanding [74]. Peter et al. showed that these patients present deficits in understanding emotions and do not adequately use their ability to regulate their emotions (that are not altered) [75]. The frequent amplification of emotional experiences would lead to frequent maladaptive responses, and the repetition of these maladaptive responses will negatively affect their social relationships. Alternatively, when BPD people inadequately use their emotion regulation strategies, depending on their motivation to care, they might alternate between the contamination path (if the motivation to care is too high) and indifference and cold attitudes (if the motivation to care is low). These alternations are both related to negative social consequences.

5. Conclusion

Our theoretical model combines several approaches to describe the empathic process [15–17,37,40,41].[4] In addition, all relationships between its specific features are explained by experimental findings and systematic reviews. Hence, we hypothesize that this model would be helpful for researchers, clinicians, and healthcare professionals aiming to study and predict empathy skills, opening new avenues of clinical applications and research. Furthermore, it was illustrated that the model offers suitable theoretical explanations of how empathy presumably works in three clinical examples. Finally, this new model (1) is inserted into the continuity of previous theories and findings, (2) helps analyze past research and clinical manifestations, and (3) can be empirically tested.

The current model also promotes several recommendations in line with these clinical and research applications. Firstly, it advocates stopping to use total empathy scores in experimental studies because cognitive and affective domains (and subdomains, like fantasy and perspective-taking) cannot be associated without losing crucial information. As displayed, they involve drastically different features interacting together. For the same reasons, experts (e.g., clinicians) should avoid using generic terms alone, like "empath" or qualifying someone as an empathic person.

Secondly, we also believe there is a strong necessity to reform the undifferentiated promotion of empathy skills. Indeed, healthcare professionals and professionals from other disciplines need tailored interventions to promote empathy skills. For this purpose, education requires models explaining how empathy works. Therefore, future healthcare professionals' education needs to focus on identification processes (i.e., between a patient and a therapist), teaching adaptive emotion regulation strategies (e.g., positive reappraisal, acceptance, or positive refocusing), promoting appropriate levels of motivation to care, and encouraging perspective-taking. In addition, education needs to identify middle and long-term adaptive responses that should be promoted and teach how to avoid maladaptive responses, like dramatizing or overreacting.

However, despite the model stemming from empirical findings, therein lies a speculative model at this stage. Furthermore, the model is centered on the empathizer and does not consider the interaction between the empathee and the empathizer, contrary to de Waal and Preston's model [16]. Like Bernard Rimé's interpersonal dynamic model of the social sharing of emotion [76,77], de Waal and Preston's model is important because it completes the model proposed in the current paper by highlighting the effects of the interpersonal dynamic. Nevertheless, to our knowledge, this is the first idiosyncratic model (centered on the empathizer) synthesizing the complexity of the empathy phenomenon by combining several approaches and providing concrete clinical and educational applications.

[4] The empathy domains proposed by Davis are illustrated in our model [15]. Empathic concern corresponds to the motivation to care (see Fig. 1); Personal distress refers to submersion; Fantasy corresponds to identification; and Perspective-taking to perspective-taking.

Disclosure of interest

The authors declare that they have no competing interest.

Fundings

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

None.

References

[1] Eklund JH, Meranius MS. Toward a consensus on the nature of empathy: a review of reviews. Patient Educ Couns 2021;104(2):300–7, http://dx.doi.org/10.1016/j.pec.2020.08.022.
[2] Kataoka H, Iwase T, Ogawa H, et al. Can communication skills training improve empathy? A six-year longitudinal study of medical students in Japan. Med Teach 2019;41(2):195–200, http://dx.doi.org/10.1080/0142159X.2018.1460657.
[3] Avlogiari E, Karagiannaki SM, Panteris E, et al. Improvement of medical students' empathy levels after an intensive experiential training on empathy skills. Psychiatry Clin Psychopharmacol 2021;31(4):392–400, http://dx.doi.org/10.5152/pcp.2021.21089.
[4] Ter Beest H, van Bemmel M, Adriaansen M. Nursing student as patient: experiential learning in a hospital simulation to improve empathy of nursing students. Scand J Caring Sci 2018;32(4):1390–7, http://dx.doi.org/10.1111/scs.12584.
[5] Ozcan CT, Öksüz E, Oflaz F. Improving empathy in nursing students: a comparative longitudinal study of two curricula. J Korean Acad Nurs 2018;48(5):497–505, http://dx.doi.org/10.4040/jkan.2018.48.5.497.
[6] Sjoquist LK, Cailor S, Conkey L, et al. A simulated patient experience to improve pharmacy student empathy for patients on hemodialysis. Curr Pharm Teach Learn 2020;12(7):827–33, http://dx.doi.org/10.1016/j.cptl.2020.02.018.
[7] Bas-Sarmiento P, Fernández-Gutiérrez M, Baena-Baños M, et al. Empathy training in health sciences: a systematic review. Nurse Educ Pract 2020;44:102739, http://dx.doi.org/10.1016/j.nepr.2020.102739.
[8] Walther J, Brewer MA, Sochacka NW, et al. Empathy and engineering formation. J Eng Educ 2020;109(1):11–33, http://dx.doi.org/10.1002/jee.20301.
[9] Ewin N, Chugh R, Muurlink O, et al. Empathy of project management students and why it matters. Procedia Comput Sci 2021;181:503–10, http://dx.doi.org/10.1016/j.procs.2021.01.196.
[10] Chen D, Lew R, Hershman W, et al. A cross-sectional measurement of medical student empathy. J Gen Intern Med 2007;22(10):1434–8, http://dx.doi.org/10.1007/s11606-007-0298-x.
[11] Hojat M, Vergare MJ, Maxwell K, et al. The devil is in the third year: a longitudinal study of erosion of empathy in medical school. Acad Med 2009;84(9):1182–91, http://dx.doi.org/10.1097/ACM.0b013e3181b17e55.
[12] Triffaux JM, Tisseron S, Nasello JA. Decline of empathy among medical students: Dehumanization or useful coping process? Encephale 2019;45:3–8, http://dx.doi.org/10.1016/j.encep.2018.05.003.
[13] Cuff BM, Brown SJ, Taylor L, et al. Empathy: a review of the concept. Emot Rev 2016;8(2):144–53 [https://doi.org/10.1177%2F1754073914558466].
[14] Triffaux J.M., Tisseron S., Nasello, J.A. Medical students' empathy during the COVID-19 pandemic: A cross-sectional study. In Press.
[15] Davis MH. A multidimensional approach to individual differences in empathy. JSAS Cat Sel Doc Psychol 1980;10:85.
[16] de Waal FB, Preston SD. Mammalian empathy: behavioural manifestations and neural basis. Nat Rev Neurosci 2017;18(8):498–509, http://dx.doi.org/10.1038/nrn.2017.72.
[17] Preston SD, de Waal FB. Empathy: its ultimate and proximate bases. Behav Brain Sci 2002;25(1):1–20, http://dx.doi.org/10.1017/S0140525X02000018.
[18] Wróbel M, Królewiak K. Do we feel the same way if we think the same way? Shared attitudes and the social induction of affect. Basic Appl Soc Psychol 2017;39(1):19–37, http://dx.doi.org/10.1080/01973533.2016.1227709.
[19] Lamm C, Decety J, Singer T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced

pain and empathy for pain. Neuroimage 2011;54(3):2492–502, http://dx.doi.org/10.1016/j.neuroimage.2010.10.014.
[20] Lockwood PL. The anatomy of empathy: vicarious experience and disorders of social cognition. Behav Brain Res 2016;311:255–66, http://dx.doi.org/10.1016/j.bbr.2016.05.048.
[21] Wicker B, Keysers C, Plailly J, et al. Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. Neuron 2003;40(3):655–64, http://dx.doi.org/10.1016/S0896-6273(03)00679-2.
[22] Preston SD. A perception-action model for empathy. In: Farrow T, Woodruff P, editors. Empathy in mental illness. Cambridge, Cambridge University Press; 2007. p. 428–47 [https://doi.org/10.1017/CBO9780511543753.024].
[23] Morelli SA, Lieberman MD. The role of automaticity and attention in neural processes underlying empathy for happiness, sadness, and anxiety. Front Hum Neurosci 2013;7:160, http://dx.doi.org/10.3389/fnhum.2013.00160.
[24] Zaki J, Kallman S, Wimmer GE, et al. Social cognition as reinforcement learning: feedback modulates emotion inference. J Cogn Neurosci 2016;28(9):1270–82, http://dx.doi.org/10.1162/jocn_a_00978.
[25] Epstude K, Mussweiler T. What you feel is how you compare: how comparisons influence the social induction of affect. Emotion 2009;9(1):1–14, http://dx.doi.org/10.1037/a0014148.
[26] Van Der Schalk J, Fischer A, Doosje B, et al. Convergent and divergent responses to emotional displays of ingroup and outgroup. Emotion 2011;11(2):286–98, http://dx.doi.org/10.1037/a0022582.
[27] Weisbuch M, Ambady N. Affective divergence: automatic responses to others' emotions depend on group membership. J Pers Soc Psychol 2008;95(5):1063–79, http://dx.doi.org/10.1037/a0011993.
[28] Yamamoto S. Primate empathy: three factors and their combinations for empathy-related phenomena. Wiley Interdiscip Rev Cogn Sci 2017;8(3):e1431, http://dx.doi.org/10.1002/wcs.1431.
[29] Singer T. The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. Neurosci Biobehav Rev 2006;30(6):855–63, http://dx.doi.org/10.1016/j.neubiorev.2006.06.011.
[30] Dixon-Gordon KL, Aldao A, De Los Reyes A. Emotion regulation in context: examining the spontaneous use of strategies across emotional intensity and type of emotion. Pers Individ Dif 2015;86:271–6, http://dx.doi.org/10.1016/j.paid.2015.06.011.
[31] McRae K, Gross JJ. Emotion regulation. Emotion 2020;20(1):1–9, http://dx.doi.org/10.1037/emo0000703.
[32] De Vaus J, Hornsey MJ, Kuppens P, et al. Exploring the East-West divide in prevalence of affective disorder: a case for cultural differences in coping with negative emotion. Pers Soc Psychol Rev 2018;22(3):285–304 [https://doi.org/10.1177%2F1088868317736222].
[33] Garnefski N, Kraaij V, Spinhoven P. Negative life events, cognitive emotion regulation and emotional problems. Pers Individ Dif 2001;30(8):1311–27, http://dx.doi.org/10.1016/S0191-8869(00)00113-6.
[34] Gross JJ. The emerging field of emotion regulation: an integrative review. Rev Gen Psychol 1998;2(3):271–99 [https://doi.org/10.1037%2F1089-2680.2.3.271].
[35] Gross JJ. Emotion regulation: current status and future prospects. Psychol Inq 2015;26(1):1–26, http://dx.doi.org/10.1080/1047840X.2014.940781.
[36] Decety J, Cowell JM. Empathy, justice, and moral behavior. AJOB Neurosci 2015;6(3):3–14, http://dx.doi.org/10.1080/21507740.2015.1047055.
[37] Decety J. The neural pathways, development and functions of empathy. Cur Op Behav Sci 2015;3:1–6, http://dx.doi.org/10.1016/j.cobeha.2014.12.001.
[38] Jonsson A, Segesten K. The meaning of traumatic events as described by nurses in ambulance service. Accid Emerg Nurs 2003;11(3):141–52, http://dx.doi.org/10.1016/S0965-2302(02)00217-5.
[39] Batson CD, Lishner DA, Stocks EL. The empathy-altruism hypothesis. In: Schroeder DA, Graziano WG, editors. The Oxford handbook of prosocial behavior. Oxford University Press; 2015. p. 259–81 [https://doi.org/10.1093/oxfordhb/9780195399813.013.023].
[40] Batson CD. Prosocial motivation: is it ever truly altruistic? Adv Exp Soc Psychol 1987;20:65–122, http://dx.doi.org/10.1016/S0065-2601(08)60412-8.
[41] Batson CD. Altruism in humans. Oxford University Press; 2011.
[42] Decety J, Bartal IBA, Uzefovsky F, et al. Empathy as a driver of prosocial behaviour: highly conserved neurobehavioural mechanisms across species. Philos Trans R Soc B Biol Sci 2016;371(1686):20150077, http://dx.doi.org/10.1098/rstb.2015.0077.
[43] Barrett LF, Mesquita B, Gendron M. Context in emotion perception. Curr Dir Psychol Sci 2011;20(5):286–90 [https://doi.org/10.1177%2F0963721411422522].
[44] Fry BN, Runyan JD. Teaching empathic concern and altruism in the smartphone age. J Moral Educ 2018;47(1):1–16, http://dx.doi.org/10.1080/03057240.2017.1374932.
[45] Nasello JA, Triffaux MS, Triffaux JM. The intergroup empathy bias among incoming medical students. Med Educ Online 2018;23:1527625, http://dx.doi.org/10.1080/10872981.2018.1527625.
[46] Preston SD, de Waal FB. Only the PAM explains the personalized nature of empathy. Nat Rev Neurosci 2017;18(12):769, http://dx.doi.org/10.1038/nrn.2017.140.
[47] Maslach C, Schaufeli WB, Leiter MP. Job burnout. Ann Rev Psychol 2001;52(1):397–422, http://dx.doi.org/10.1146/annurev.psych.52.1.397.
[48] Maslach C, Jackson SE, Leiter MP. MBI: Maslach burnout inventory. Palo Alto, CA: Consulting Psychologists Press; 1996.
[49] von Harscher H, Desmarais N, Dollinger R, et al. The impact of empathy on burnout in medical students: new findings. Psychol Health Med 2018;23(3):295–303, http://dx.doi.org/10.1080/13548506.2017.1374545.
[50] Wilkinson H, Whittington R, Perry L, et al. Examining the relationship between burnout and empathy in healthcare professionals: a systematic review. Burn Res 2017;6:18–29, http://dx.doi.org/10.1016/j.burn.2017.06.003.
[51] Kleynhans R, Markham L, Meyer W, et al. Human resource management fresh perspectives. South Africa: Person Prentice Hall; 2006.
[52] Byrne BM. Burnout: testing for the validity, replication, and invariance of causal structure across elementary, intermediate, and secondary teachers. Am Educ Res J 1994;31(3):645–73 [https://doi.org/10.3102%2F00028312031003645].
[53] Lee M, Jang KS. Nurses' emotions, emotion regulation and emotional exhaustion. International Int J Organ Anal 2019;27(5):1409–21, http://dx.doi.org/10.1108/IJOA-06-2018-1452.
[54] Herberman Mash HB, Ursano RJ, Benevides KN, et al. Identification with terrorist victims of the Washington, DC sniper attacks: posttraumatic stress and depression. J Trauma Stress 2016;29(1):41–8, http://dx.doi.org/10.1002/jts.22069.
[55] Nagamine M, Shigemura J, Fujiwara T, et al. The relationship between dispositional empathy, psychological distress, and posttraumatic stress responses among Japanese uniformed disaster workers: a cross-sectional study. BMC Psychiatry 2018;18(1):1–7, http://dx.doi.org/10.1186/s12888-018-1915-4.
[56] Hare RD. The psychopathy checklist – Revised. Toronto: Multi-Health Systems, Incorporated; 2003.
[57] Blair RJR. Fine cuts of empathy and the amygdala: dissociable deficits in psychopathy and autism. Q J Exp Psychol 2008;61(1):157–70 [https://doi.org/10.1080%2F17470210701508855].
[58] Dolan M, Fullam R. Theory of mind and mentalizing ability in antisocial personality disorders with and without psychopathy. Psychol Med 2004;34(6):1093–102, http://dx.doi.org/10.1017/S003329170400202.
[59] Seara-Cardoso A, Neumann C, Roiser J, et al. Investigating associations between empathy, morality and psychopathic personality traits in the general population. Pers Indiv Dif 2012;52(1):67–71, http://dx.doi.org/10.1016/j.paid.2011.08.029.
[60] Van Dongen JD. The empathic brain of psychopaths: from social science to neuroscience in empathy. Front Psychol 2020;11:695, http://dx.doi.org/10.3389/fpsyg.2020.00695.
[61] Burghart M, Mier D. No feelings for me, no feelings for you: a meta-analysis on alexithymia and empathy in psychopathy. Pers Indiv Dif 2022;194:111658, http://dx.doi.org/10.1016/j.paid.2022.111658.
[62] Hare RD. Psychopathy: a clinical construct whose time has come. Crim Justice Behav 1996;23(1):25–54 [https://doi.org/10.1177%2F0093854896023001004].
[63] Hare RD, Neumann CS. Psychopathy: assessment and forensic implications. Can J Psyc 2009;54(12):791–802 [https://doi.org/10.1177%2F070674370905401202].
[64] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. (DSM-5). Washington: American Psychiatric Publishing; 2013.
[65] Martin F, Flasbeck V, Brown EC, et al. Altered mu-rhythm suppression in borderline personality disorder. Brain Res 2017;1659:64–70, http://dx.doi.org/10.1016/j.brainres.2017.01.023.
[66] New AS, Carpenter DM, Perez-Rodriguez MM, et al. Developmental differences in diffusion tensor imaging parameters in borderline personality disorder. J Psychiatr Res 2013;47(8):1101–9, http://dx.doi.org/10.1016/j.jpsychires.2013.03.021.
[67] Petersen R, Brakoulias V, Langdon R. An experimental investigation of mentalization ability in borderline personality disorder. Compr Psychiatry 2016;64:12–21, http://dx.doi.org/10.1016/j.comppsych.2015.10.004.
[68] Flasbeck V, Enzi B, Brüne M. Childhood trauma affects processing of social interaction in borderline personality disorder: an event-related potential study investigating empathy for pain. World J Biol Psychiatry 2019;20(4):278–88, http://dx.doi.org/10.1080/15622975.2017.1333147.
[69] Grzegorzewski P, Kulesza M, Pluta A, et al. Assessing self-reported empathy and altruism in patients suffering from enduring borderline personality disorder. Psychiatry Res 2019;273:798–807, http://dx.doi.org/10.1016/j.psychres.2018.12.109.
[70] Harari H, Shamay-Tsoory SG, Ravid M, et al. Double dissociation between cognitive and affective empathy in borderline personality disorder. Psychiatry Res 2010;175(3):277–9, http://dx.doi.org/10.1016/j.psychres.2009.03.002.
[71] Bujalski AS, Chesin MS, Jeglic EL. Cognitive empathy partially mediates the relationship between childhood physical abuse and borderline personality disorder features in college students. Curr Psychol 2019;38(1):121–7, http://dx.doi.org/10.1007/s12144-017-9597-5.
[72] Matzke B, Herpertz SC, Berger C, et al. Facial reactions during emotion recognition in borderline personality disorder: a facial electromyography study. Psychopathology 2014;47(2):101–10, http://dx.doi.org/10.1159/000351122.
[73] Dziobek I, Fleck S, Kalbe E, et al. Introducing MASC: a movie for the assessment of social cognition. J Autism Dev Disord 2006;36(5):623–36, http://dx.doi.org/10.1007/s10803-006-0107-0.
[74] Gardner K, Qualter P. Emotional intelligence and borderline personality disorder. Pers Individ Dif 2009;47(2):94–8, http://dx.doi.org/10.1016/j.paid.2009.02.004.

[75] Peter M, Arntz AR, Klimstra T, et al. Different aspects of emotional intelligence of borderline personality disorder. Clin Psychol Psychother 2017;25(1):e51–9, http://dx.doi.org/10.1002/cpp.2126.
[76] Rimé B. Le partage social des émotions. Presses universitaires de France; 2005.
[77] Rimé B. Emotion elicits the social sharing of emotion: theory and empirical review. Emot Rev 2009;1(1):60–85 [https://doi.org/10.1177%2F1754073908097189].
[78] Decety J. Composants, mécanismes, développement et fonctions de l'empathie. EMC Psychiatrie 2016;13(1):1–8.
[79] Chismar D. Empathy and sympathy: the important difference. J Value Inq 1988;22(4):257–66.
[80] Bruneau EG, Cikara M, Saxe R. Parochial empathy predicts reduced altruism and the endorsement of passive harm. Soc Psychol Personal Sci 2017;8(8):934–42 [https://doi.org/10.1177%2F1948550617693064].
[81] Decety J. Why empathy is not a reliable source of information in moral decision making? Curr Dir Psychol Sci 2021;30(5):425–30 [https://doi.org/10.1177%2F09637214211031943].
[82] Fourie MM, Subramoney S, Gobodo-Madikizela P. A less attractive feature of empathy: intergroup empathy bias. In: Kondo M, editor. Empathy: an evidence-based interdisciplinary perspective. INTECH; 2017. p. 45–61.

Parallèle entre les conceptions de l'empathie et le DPM

Le modèle présenté a permis d'illustrer la complexité de l'empathie en montrant l'articulation dynamique des différentes facettes. Toutefois, des auteurs (Yu & Chou, 2018) ont proposé une approche nettement simplifiée du phénomène en décomposant celui-ci en deux voies principales : (1) une voie rapide qui est propre à la facette affective de l'empathie qui va impliquer le système des neurones miroirs et (2) une voie lente, propre à la facette cognitive de l'empathie, qui va engager des régions cérébrales telles que le précunéus, la jonction temporopariétale et le cortex préfrontal médian, moyen et dorsal.

Cette simplification du phénomène de l'empathie proposée par Yu et Chou (2018) a dû naturellement évoquer un autre modèle que nous avons abordé précédemment, le DPM. Effectivement, le parallélisme entre les deux modèles est assez flagrant, ce qui permet d'utiliser l'empathie comme un phénomène de prédilection pour tester la validité du DPM étant donné qu'elle fait intervenir les mêmes composantes.

Le présent travail et son décours

A partir de ce cadre théorique, le présent travail s'est développé sur les bases suivantes : (1) le développement de dilemmes de type trolley dans des situations plus écologiques (ce que l'on a nommé des *everyday trolley-like problems*) et la comparaison de ces dilemmes avec les dilemmes classiques (études 2 & 3), (2) l'utilisation du DPM (Greene et al., 2001) comme outil de prédiction des décisions morales, (3) la sélection de l'empathie, un phénomène impliquant des composantes cognitives et émotionnelles, comme potentiel mécanisme explicatif des prises de décision morales et, enfin, (4) le fait de tester ces différents dilemmes sur une population psychopathologique présentant un trouble de la personnalité borderline.

Dès lors, nous avons réalisé : (1) trois études majeures (deux portant sur la population générale et une portant sur une population clinique d'individus présentant un trouble de personnalité borderline), (2) traduit une échelle évaluant les traits de personnalité borderline en français, (3) développé un modèle théorique de l'empathie et ajusté un modèle de prise de décision et, enfin, (4) produit une revue systématique et méta-analyse de la littérature scientifique afin de clarifier les liens entre les facettes de l'empathie et les jugements moraux, les prises de décision morales et l'inclinaison morale.

Section 2 : Partie expérimentale

Chapitre 2 : première étude

Etude 1 : étude des liens entre empathie, perspective et choix moraux dans des dilemmes de type trolley

Notre première étude visait à tester les points suivants : l'empathie comme caractéristique prédictive des choix moraux dans des dilemmes de type trolley de la vie de tous les jours. Nous avons dans un premier temps créé un dilemme de type trolley qui impliquait que le (la) participant(e) incarne un (e) conducteur(trice)[3] de tram allant d'une gare à une université, sans arrêt intermédiaire (le trajet complet dure 1h30). En bref, ce tram est paramétré en fonction des horaires de cours et la politique de l'université est particulièrement stricte sur les retards. En effet, certains étudiants peuvent se voir refuser l'accès à leur examen de première session s'ils se présentent en retard. Bien engagé dans sa course, le (la) conducteur(trice) reçoit un appel de sa centrale lui expliquant qu'un(e) étudiant(e)[4] se trouve sur une station au milieu du trajet. Son tram étant complet et n'étant pas censé s'arrêter, la centrale explique que le(la) conducteur(trice) peut prendre la décision qu'il(elle) souhaite : s'arrêter pour charger l'étudiant(e) retardataire ou continuer sa course. S'il(elle) s'arrête, tout le monde devra se serrer davantage et on estime un retard de 15 minutes. Dès lors, certains étudiants, déjà présents dans le tram, arriveront en retard à leurs cours et certains pourraient donc se voir refuser l'accès à leur examen. Toutefois, le(a) retardataire accédera à son cours de manière certaine sans pénalité. Si le(a) conducteur(trice) ne s'arrête pas, tous les étudiants déjà présents arriveront à leurs cours dans les temps (voir Annexe A pour une description détaillée).

Avec ce matériel, nous avons créé deux conditions : une où le(a) participant(e) incarnait le(a) conducteur(trice) et une où ce(tte) dernier(ière) incarnait le(a) retardataire. Les deux conditions étaient complétées aléatoirement pour chaque participant(e). Ici, un choix utilitaire correspondait au fait de ne pas s'arrêter lorsque le(a) sujet incarnait le(a)

[3] Dans le but de rendre les scénarios plus immersifs, ils étaient attribués comme suit : une femme incarnait une conductrice et un homme, un conducteur. De plus, afin de maximiser un effet de l'empathie, le(a) retardataire étaient du même sexe et même âge que le(a) participant(e). En effet, d'après Preston et de Waal (2002), cinq caractéristiques jouent un rôle prépondérant dans l'empathie : la similarité (i.e., le rapprochement perçu entre un sujet et un objet comme la personnalité, le genre ou l'âge), la familiarité, l'apprentissage, l'expérience passée et la saillance.

[4] Il est mentionné que le(a) retardataire a une raison considérée comme valable par le(a) participant(e) d'être arrivé(e) en retard.

conducteur(trice) ou de ne pas faire arrêter le tram lorsque ce(tte) dernier(ière) incarnait le(a) retardataire.

Nous avons sélectionné la *Basic Empathy Scale* (Jolliffe & Farrigton, 2006 ; version française : D'Ambrosio et al., 2009) comme outil de mesure de l'empathie car celle-ci est principalement focalisée sur les émotions (voir Annexe B), contrairement à son homologue plus populaire l'*Interpersonal Reactivity Index* (Davis, 1980 ; Davis, 1983) qui étudie l'empathie selon tous les états mentaux confondus (i.e., pensées, sentiments, comportements). Une comparaison détaillée est réalisée dans l'article qui va suivre (Nasello et al., 2020).

Les hypothèses de cette première étude ont été formulées à partir du DPM et de publications précédentes ayant démontré des liens significatifs entre l'empathie et les prises de décision morale (Gleichgerrcht & Young, 2013 ; Patil & Silani, 2014). Ces dernières ont montré que la préoccupation empathique (une facette de l'empathie affective) était le principal prédicteur des choix moraux. Dès lors, sur la base du DPM, nous nous attendions à ce qu'une personne ayant un faible niveau d'empathie affective soit plus encline à faire des choix utilitaires, surtout dans la situation où elle incarne le(a) conducteur(trice) (situation comparable à un dilemme de type *impersonnel*). En revanche, un haut niveau d'empathie affective devrait conduire cette personne à réaliser moins de choix utilitaires, surtout dans la situation où elle incarne le(a) retardataire (situation comparable à un dilemme de type *personnel*).

En résumé, une association négative était attendue entre l'empathie affective et les choix utilitaires. De plus, certains auteurs ont rapporté que les hommes font davantage de choix utilitaires que les femmes et nous nous attendions à observer cette même différence (Fumagalli et al., 2010a ; Fumagalli et al., 2010b ; Armstrong et al., 2019).

Article : Does empathy predict decision-making in everyday trolley-like problems?

Current Psychology
https://doi.org/10.1007/s12144-021-01566-1

Does empathy predict decision-making in everyday trolley-like problems?

Julian A. Nasello[1,2] · Benoit Dardenne[3] · Adélaïde Blavier[2] · Jean-Marc Triffaux[1,4]

Accepted: 2 March 2021

Abstract

Two ecological and social versions of the Trolley-like Problem (a *direct-impact* [DI] and a *no-direct-impact* [NDI] scenario) were created to investigate moral decision-making and determine whether age, gender, and empathy significantly predicted people's choices. Two hundred and fifty-five participants were recruited in this study. We found that men and women made more *Utilitarian* Choices (UC) in the NDI scenario. In this scenario, the attribution of penalties was low (6%). In the DI scenario, we found that the UC rate was significantly reduced compared to the NDI scenario, and, surprisingly, women made more UC than men. We found that UC decreased as a function of men's affective empathy in the NDI scenario. In the DI one, lower affective empathy levels predicted UC differently for men and women. Lastly, the attribution of penalties tripled. In conclusion, this study highlighted the role played by scenarios, gender, and empathy in predicting UC in Trolley-like Problems.

Keywords Utilitarianism · Moral decision-making · Trolley problems · Social dilemmas · Empathy

Introduction

For decades, social psychology and economics have investigated *utilitarianism* (i.e., the morally correct course of action producing the most significant total benefit for all people affected) in social dilemma situations (i.e., social situations involving a moral conflict). Classically, the authors have studied *utilitarian* and *non-utilitarian* (also called *deontological*) decisions by using Trolley Problems and other variants. Greene et al. (2001) have popularized these problems and, initially, have focused their attention on two different Trolley Problems: the Foot (1967) and Thomson's (1976) problems. The former was qualified as *impersonal* and the latter as *personal*.[1]

[1] For a review of these scenarios, see Bruers and Braeckman (2014).

✉ Julian A. Nasello
julian.nasello@gmail.com

[1] Psychosomatic Medicine and Group Psychotherapy, University Day Hospital "La Clé", Boulevard de la Constitution 153, 4020 Liège, Belgium

[2] Department of Clinical Psychology, University of Liège, Liège, Belgium

[3] Department of Social Psychology, University of Liège, Liège, Belgium

[4] Department of Psychiatry, Medicine, University of Liège, CHU of Liège, Liège, Belgium

Foot's (1967) *impersonal* scenario is usually presented as follows: *"*[the participant] *is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed.*" (Foot 1967 p. 2). The dilemma consists of a choice between sacrificing one worker and saving the five workers on the other track, or sacrificing the five, and saving the one. According to Greene et al. (2001, 2004; Greene and Haidt 2002; Greene 2007) and Koenigs et al. (2007), a *utilitarian* moral decision in the Trolley Problem consists of choosing to hit one individual to save the other five through a cost-benefit analysis that will produce a rational response aiming to maximize aggregate welfare. On the other hand, a *non-utilitarian moral* decision (also called *deontological*) consists of allowing the Trolley to hit the five to save the one (Gleichgerrcht and Young 2013). In contrast, Thomson's (1976) *personal* scenario (also called the *footbridge* Trolley Problem) puts the participant on a footbridge. S/he sees an empty runaway trolley speeding down a set of tracks toward five railway workers. If s/he does nothing, these five men are going to be killed by the Trolley. Next to him/her, another obese workman is standing, and pushing him will stop the Trolley and save the five workers.

Greene et al. (2001) have argued that *personal* moral decisions are driven mostly by social-emotional responses (e.g., Thomson's Trolley problem), while *impersonal* moral decisions are driven less by social-emotional responses and more

Published online: 25 March 2021

by cognitive processes (e.g., Foot's Trolley Problem). Accordingly, they proposed a dual-process morality theory where one system is fast, intuitive, and mostly involves emotional processes, whereas the other one is slow, elaborative, and mostly involves cognitive processes. The former produces higher rates of *non-utilitarian* decisions (e.g., Copraro and Sippel 2017; Gleichgerrcht and Young 2013) and the later higher *utilitarian* decision rates (e.g., Navarrete et al. 2012; Gleichgerrcht and Young 2013).

Scientific research classically investigates whether men and women present significant differences in Trolley's decision-making. Authors found that men showed more *utilitarian* choices than women in *personal* (i.e., Thomson's Trolley Problem) and more emotionally salient dilemmas (Fumagalli et al. 2010a; Fumagalli et al. 2010b). Besides, no gender difference was found in *impersonal* dilemmas. Recent meta-analyses confirmed this slight gender effect, showing that men present more *utilitarian* inclinations than women, and women more *deontological* inclination than men (Friesdorf et al. 2015; Armstrong et al. 2019). These inclinations to display more *deontological* judgments for women are presumably due to stronger affective responses in harming someone; while men are engaged in more cognitive evaluations of outcomes than affective ones.

Relations between *utilitarianism* and Empathy

Ramsøy et al. (2015 p. 179) stated that "*the decision whether or not to act socially is a conflict between what is optimal for the individual, as opposed to what is optimal for the society* [or the group] *at large*." To highlight the complexity of factors that lead people to make decisions, some authors focused on the role played by empathy (Davis 2015), notably because it is a prosocial feeling that causes people to help others (Batson et al. 1997; de Waal and Suchak 2010). Empathy is an essential aspect of moral judgment because understanding and sharing emotions with others may discourage people from moral transgressions that cause suffering (Vyas et al. 2017).

Empathy has been defined as "*any process where the attended perception of the object's state generates a state in the subject that is more applicable to the object's state or situation than to the subject's own prior state or situation*" (Hoffman 2000), and also as "*the innate ability to perceive and be sensitive to the emotional states of others coupled with a motivation to care for their well-being.*" (Decety 2015). These definitions showed that empathy is both a dispositional trait and a process. Jean Decety (2015 p.1) divided empathy into three domains: 1) affective sharing, defined as "*the capacity to share or become affectively aroused by others' emotional valence and relative intensity without confusion between self and other*"; 2) empathic concern, described as "*the motivation to care for another's welfare*"; and 3) perspective-taking (or cognitive empathy), which refers to "*the ability to consciously put oneself into the mind of another and understand what that person is thinking or feeling.*"

Several authors were intrigued by the significant contribution of empathy in *utilitarian* decision-making. Indeed, they found that the lack of dispositional empathic concern appeared to be a significant predictor of more *utilitarian* decisions in *personal* and *impersonal* scenarios; and that neither personal distress, nor perspective-taking, nor fantasy (i.e., three other domains of empathy) were consistent predictors for such decision (Gleichgerrcht and Young 2013; Decety and Cowell 2014; Patil and Silani 2014). These results are in line with the empathy-altruism hypothesis (Batson et al. 2015) that states that "*empathic concern produces altruistic motivation*" (Batson et al. 2015 p. 1).

As mentioned earlier, Greene et al. (2001, 2004; Greene and Haidt 2002; Greene 2007) proposed a Dual-Process Model explaining how people make *utilitarian* and *deontological* decisions in Trolley Problems (i.e., *personal* and *impersonal*[2]). According to people's affective empathy levels, this model can be extended by making four predictions (see Table A in Supplementary Materials): individuals with higher affective empathy levels would experience higher emotions and be less involved in reasoning processes that engage a cost-benefit analysis. Therefore, they will produce lower rates of utilitarian choices in *impersonal* scenarios (prediction 1) and even lower ones in *personal* scenarios (prediction 2). In contrast, individuals low in affective empathy would experience lower emotions andbe highly involved in cognitive processes. They will make higher utilitarian choices in *personal* scenarios (prediction 3) and even higher utilitarian choices in *impersonal* scenarios (prediction 4).

Current Limits of Trolley Problems and Solutions

Despite the interesting influence of affective empathy on moral decision-making, some authors advised readers to be wary of the claim that the Trolley Problems can assess morality (Bauman et al. 2014) and *utilitarian* decision-making (Kahane 2015).

According to Bauman et al. (2014), the Trolley problems are neither immersive nor realistic: participants tend to be amused rather than concerned about the situation that gives

[2] *Impersonal* and *personal* classification can be transposed in the scenarios used in this study. As readers will see in the Materials section, Scenario 1 is comparable to *impersonal* dilemmas because the participant is not invited to violate his/her moral by giving intentional harm to someone. We considered this scenario as a *no-direct-impact* scenario because the consequences of the participant's choice will not impact him/her directly. In this kind of scenario, lower levels of emotions are elicited and participants are more inclined to get involved in cognitive processes to make a pragmatic choice. Moreover, we compare *personal* scenarios with Scenario 2 because participant's choices have *direct-impact* on him/her (see Materials section for further details). In this scenario, higher levels of emotions are elicited, leaving the participant less involved in cognitive processes.

rise to the dilemma. Thus, in its classical form, the Trolley problems are too far from a real-world situation. Lastly, the psychological processes involved in such an unrealistic situation may not be the same as in the "real world." To mitigate these limitations and make the Trolley problems more immersive, some authors have implemented these Problems in Virtual Reality (Navarrete et al. 2012). Surprisingly, results were in line with those from the "paper format": participants made more *utilitarian* choices (around 90%) in both *omission* and *action* conditions (i.e., in the *action* condition, participants had to pull the switch to kill one to save the five; in the *omission* condition, participants had to do nothing to kill one to save the five). However, recent findings showed that hypothetical dilemmas have little predictive value for actual behaviors (Bostyn et al. 2018). The authors provided evidence that the participants' preference for utilitarian decision-making was not related to their decision in a real-life dilemma. Their judgments in the hypothetical scenarios significantly predicted their *lack of doubt* (i.e., participants showing preferences for utilitarian reasoning made a speedier utilitarian judgment) and the *degree of discomfort felt* in the real-life dilemma (i.e., participants showing preferences for utilitarian reasoning felt less discomfort in the real-life scenario). The authors concluded that "*important aspects of the real-life decision-making process are not captured through the standard trolley paradigm*." (Bostyn et al. 2018, p. 1090). Therefore, there is a crucial need to improve this kind of material. For instance, FeldmanHall et al. (2012) stated that "*the underspecified and impoverished nature of hypothetical moral probes is unable to capture the complex social, emotional, and motivational pressures inherent to real moral decisions*." (p. 440), empowering the necessity to use more realistic materials to catch moral decision-making, like everyday Trolley-like Problems.

The *action* or *omission* of killing someone (i.e., an absolute consequence) is, fortunately, far from daily-life situations. Indeed, some authors consider that the classical Trolley Problems might be likely to elicit more *utilitarian* decision-making than in everyday dilemmas (Vyas et al. 2017). Recent studies used adaptations of the Trolley Problems to everyday life situations (Lotto et al. 2014; Vyas et al. 2017; Pletti et al. 2017; Takamatsu 2019) and showed more generalizable findings. The use and development of ecological Trolley-like Problems appear as a priority, notably because there is a growing body of research investigating moral decisions in several psychopathologies (e.g., psychopathy (Koenigs et al. 2012; Gao and Tang 2013; Patil 2015; Balash and Falkenbach 2018); autism (Gleichgerrcht et al. 2013; Li and Liu 2017); antisocial personality disorders (Bartels and Pizarro 2011); Dark Triads (Djeriouat and Trémolière 2014); or sadism (Trémolière and Djeriouat 2016)). It will allow scientists to provide more robust evidence of factors (e.g., empathy) affecting or not moral decision-making in various clinical populations.

A last but not least limit is the nomenclature used to describe people's moral decisions in Trolley Problems. Concerning the classical taxonomy, authors have separated *utilitarian* from *non-utilitarian* choices (Greene et al. 2001). However, some authors have questioned this dichotomous classification (Kahane 2015; Kahane et al. 2015) and defended the idea that Trolley Problems tell us little about *utilitarian* or *deontological* decision-making. For instance, Kahane et al. (2015, p. 206) found that *a tendency to endorse the violent sacrifice of one person in order to save a greater number was not (or even negatively) associated with paradigmatic markers of utilitarian concern for the greater good*, meaning that deciding to save the higher number of people is not *de facto* associated with utilitarianism. Therefore, except if it is explicitly studied, we cannot know if a participant pursues maximizing others' welfare or minimizing losses. Indeed, other motives might have fostered this kind of decision, such as avoiding specific emotions (e.g., anger or guilt).

Not only it seems there is little or no association between *utilitarianism* and Trolley Problems, but also the high diversity of decision-making in these Problems can hardly stand in a dichotomic classification (*utilitarian* or *deontological* choices). For example, is it a comparable *non-utilitarian* decision between killing the five to save one and killing the five to save me? According to the classical nomenclature, if the participant's decision is not maximizing the aggregate welfare, then we describe this choice as *non-utilitarian*. However, we can see that there is a massive difference between saving one or saving me. This confusion lies in the fact that the classical definition of *non-utilitarian* decisions is apophatic (i.e., defining something by stating which characteristics it does not have). To solve this problem, we suggest a cataphatic definition of participants' decision-making (i.e., defining something by describing which features it does have). In social dilemmas, most decisions are oriented to an *Object*: to a group, a dyad, an isolated individual, or even to oneself. Therefore, it would be more appropriate to reframe moral decisions by describing to whom this decision is oriented: when a participant makes so-called *utilitarian* decisions (e.g., saving the five), we propose to describe it as "*group*-oriented choices"; when individuals favor themselves, it will be described as "*self*-oriented choices"; and when s/he makes decisions in favor of an isolated individual, then it will be described as "*individual*-oriented choices." Following the confusion that might underlie *utilitarian* or *deontological* terminology, we believe that this nomenclature will provide a more descriptive and robust language to the scientific community.

Objectives

We have created two Trolley-like scenarios for students (a *direct-impact* and a *no-direct-impact* scenario) and focused our attention on the predictive role of empathy - assessed by the Basic Empathy Scale (BES; Jolliffe and Farrington 2006) - in *utilitarian* choices.

As mentioned, we used Trolley-like Problems that do not involve harming or killing someone to make the scenarios more familiar or closer to daily-life situations. We tested participants' decisions in Scenario 1 (i.e., the *no-direct-impact* scenario where participants have to make *individual*-oriented versus *group*-oriented choices) and 2 (i.e., the *direct-impact* scenario where participants have to make *self*-oriented versus *group*-oriented choices). According to the scientific literature (Di Nucci 2013; Swann et al. 2010) and the predictions made by the Dual-Process Model (Greene et al. 2001), people make more *utilitarian* choices (i.e., *group*-oriented choices) in *impersonal* scenarios, and men tend to display more *utilitarian* choices than women in *personal* scenarios (Fumagalli et al. 2010a; Fumagalli et al. 2010b; Friesdorf et al. 2015; Armstrong et al. 2019). Therefore, we expect to observe higher rates of *group*-oriented choices than *individual*-oriented choices in the *no-direct-impact* scenario (Hypothesis 1a) and higher rates of *group*-oriented choices in the *direct-impact* scenario for men compared to women (Hypothesis 1b).

As peripheral objectives, we explore whether the participants' degree of penalty allotment (see next section) differs in Scenarios 1 and 2 (Hypothesis 2a) and whether empathy significantly predicts penalty allotment (Hypothesis 2b).

The main objective was to test the effect of empathy in participants' decision-making using more familiar Trolley-like Problems. Authors are quite unanimous in pointing out that emotions play an essential role in moral judgments (Greene et al. 2001; Greene and Haidt 2002; Greene et al. 2004; Choe and Min 2011), and we selected the Basic Empathy Scale (BES: Jolliffe and Farrington 2006) on this purpose. Hence, we expect to replicate affective empathy's significant role in so-called *group*-oriented decisions using everyday Trolley-like Problems (Hypothesis 3). Moral values are based on emotional responses (Decety and Cowell 2014; Patil and Silani 2014; Gleichgerrcht and Young 2013), and the lack of empathic concern (affective empathy[3]) showed a significant influence on *group*-oriented (*utilitarian*) moral decisions.

Materials & Methods

Participants

Initially, two hundred and sixty-seven students participated in our study voluntarily. The sample size recommendation is 267, with a margin error of 5%, a confidence level of 90%, and a response distribution of 50% (Raosoft 2004). However, following a Boxplot analysis, twelve outlier participants were excluded due to their atypical scores on affective and cognitive empathy.[4] The final sample was composed of two hundred and fifty-five participants ($n_{women} = 189$; $M_{age} = 20.9$; $SD_{age} = 2.9$; Cf. Table 1[5]). With a sample size of 255, our margin of error is 5.12%. All participants were recruited on the campus of the University of Liège (Belgium). They came from six different departments: 32% ($n = 82$) were students from the social sciences and psychology; 20% ($n = 51$) were students in medical and veterinarian school; 19% ($n = 48$) were law students; 12% ($n = 30$) were business students; 10% ($n = 25$) were students in philosophy; and 7% ($n = 19$) came from other departments (i.e., applied sciences or agro-biological technologies). Inclusion criteria consisted of fluency in French, active enrollment in University classes, a minimum age of 18 years old and maximum age of 35.

Procedure

The Faculty of Psychology's Research Ethics Committee at the University of Liège (Belgium) approved this study, Ref.: 1718-12. Students were asked to freely and anonymously participate in our research by using an online survey. They all received detailed information and gave their consent by clicking on a link provided to them. Subsequently, participants completed different questionnaires and made their choices according to the two Trolley-like Problems (distributed in counterbalanced order) described below.

Materials

Demographic information: the student participants were asked to provide their age, their gender, and their department.

Empathy: we used the Basic Empathy Scale to assess two facets of empathy (BES: Jolliffe and Farrington 2006; French version: D'Ambrosio et al. 2009; Carré et al. 2013). Jolliffe and Farrington (2006) initially developed the BES to overcome specific weaknesses of the Interpersonal Reactivity Index (IRI: Davis 1983) (*for a detailed description, see* Jolliffe and Farrington 2006; Carré et al. 2013). This scale assesses two empathy facets: affective and cognitive empathy (i.e., the ability to feel and understand others' feelings) and four basic emotions (i.e., anger, fear, happiness, and sadness). Therefore, in contrast with the IRI, all BES items are limited to emotional abilities and not to broader abilities (e.g., the IRI presents items such as understanding the other person's viewpoint, even when emotions are not involved). The BES authors defined affective empathy according to Bryant's definition (Bryant 1982) as "*the capacity to experience the emotions*

[3] It is important to note that affective empathy from the BES (Jolliffe and Farrington 2006) is significantly linked with the components of affective empathy from the Interpersonal Reactivity Index (Davis 1983). The authors showed a strong correlation of .74 between those two components (Vachon and Lynam 2016). Therefore, we can assume that empathic concern from the IRI and the affective empathy component of the BES are strongly related.

[4] Removing these outlier participants did not affect the significance of our analyses.

[5] Find the raw data file in the following link: DOI: 10.17605/OSF.IO/5Q6XN.

of another." and cognitive empathy was defined according to Hogan's (1969) formulation as "*the capacity to comprehend the emotions of another*."

Affective empathy is composed of 11 items, such as "*I get caught up in other people's feelings.",* while cognitive empathy is composed of 9 items, such as "*When someone is feeling 'down,' I can usually understand how they feel."* The two factors are evaluated on a 5-point Likert scale from 1 (*It does not describe me very well*) to 5 (*It describes me very well*). Total scores are calculated for each dimension. Note that eight items are negatively worded. The BES scale presented good internal reliability in our sample: the Cronbach's alphas for total items was .84, .85 for affective empathy, and .75 for cognitive empathy.

Social dilemmas (see Supplementary Materials): drawing on the original Trolley problems (Foot 1967; Thomson 1976), we used two different scenarios containing two different perspectives and provided these to the participants at random. Both scenarios had the same parameters: the tramway transports students to their University; the trip is long, requiring one and a half hours to get from departure to arrival (i.e., the University); the tramway does not stop during the trip; the tramway schedule is aligned with the class schedule; there are no other means of transport available; class attendance is compulsory for all students, and the University has a strict policy on lateness: in case of lateness to class, a student runs the risk of not being allowed to sit for the first year-end final exam session.[6] In Scenario 1 (the *no-direct-impact* scenario), the participant was invited to be a tramway driver who had to choose whether or not to stop the tramway for a tardy student. His/her head office contacts the driver to explain the situation and the tardy student presents a valid reason for not having boarded the tramway at the departure. However, choosing to stop the tramway is not without consequences. Indeed, the time delay is estimated at fifteen minutes, and there is a high risk that some students already present in the tramway will be late for their lecture (i.e., those whose classrooms are located the furthest away). These students, therefore, might not be allowed to sit for the first year-end final exam session. On the other hand, the tardy student is sure to attend his/her class. If the driver chooses not to stop, all students in the tramway will arrive at their lectures on time, but the tardy student will not. In this scenario, we used a generic stranger. In its conventional implementation, this generic stranger is differentiated from other people by his/her state of need, and little additional information was customarily provided. The generic stranger used in this study was a student of the same age and gender as the participant. In this way, this stranger is likely to be perceived as somewhat similar to the participant (Davis 2015).

In Scenario 2 (the *direct-impact* scenario), the same parameters are present. However, the participant is invited here to be the tardy student in the middle of the trip. As the tardy student, s/he has contacted the head office of the tramway to explain his/her situation, and this head office offers him/her the opportunity to choose whether or not s/he will stop the tramway to board.

After each scenario, we asked an additional question to assess whether the participant would be inclined to impose penalties on the tardy student (Scenario 1) or himself/herself (Scenario 2). The participant had the option to impose the following penalties on the late student or himself/herself (in the study, we have presented these penalties as "additional conditions"): 1) payment of additional fees to take the tramway; 2) deposit her/his school bag at the luggage locker of the station platform (thus making it difficult for the student to return to the mid-trip platform after class). This second question gave the following possibilities: 1) taking the tardy student or taking the tramway with additional conditions; 2) taking the tardy student or take the tramway without additional conditions; 3) not taking the tardy student or not taking the tramway. It is important to note that Scenario 1 was not crossed gender for participants: the tardy student (the generic stranger) was female for all female participants and male for all male participants.

To summarise, each participant received Scenarios 1 and 2 in random order. After reading each scenario, they had to choose whether or not to stop the tramway to take the tardy student (Scenario 1) or whether or not to stop the tramway to board themselves (Scenario 2). After this question, they can

Table 1 Descriptive statistics

Name	N	Mean	SD	Median	Min	Max
Age	255	20.9	2.90	20	18	34
Women	189	20.7	2.46	20	18	33
Men	66	21.7	3.82	20.5	18	34
AE	255	41.9	6.97	42	25	55
Women	189	43.2	6.75	43	25	55
Men	66	38.4	6.38	39	26	54
CE	255	36.7	4.05	37	27	45
Women	189	36.9	3.91	37	27	45
Men	66	36.2	4.42	36	27	45

Note. This table displays all means, standard deviations, medians, minimums, and maximums of age and empathy components (AE = Affective Empathy) and CE = Cognitive Empathy) for the whole sample, and women and men

[6] Note that at the University of Liège (Belgium), students have only two main sessions for all examinations. Missing one session puts students in the precarious position of having only one opportunity to score a passing grade. Therefore, the stakes are high if a student misses a class.

decide to impose penalties on the tardy students or themselves.

Data Analyses and Experimental Design

As descriptive statistics, we performed *Chi-square* tests to describe participants' choices in Scenarios 1 and 2. The core analysis was a Generalized Mixed Model to measure empathy's influence on the participants' decisions in Scenarios 1 and 2. For this analysis, the dependent variable was the participants' choices to stop or not ('STOP' variable) in Scenarios 1 and 2. The variable 'STOP' was coded in a binary way for both scenarios: "1″ when participants chose to stop the tramway, and "0″ when they did not stop it. For a complete description of the Generalized Mixed Model procedure, see Gallucci (2019). Age, gender (initially coded "1″ for women and "0″ for men, but Cf. footnote[7]), and BES components (affective [AE] and cognitive empathy [CE]). Note that all variables were mean-centered, and all analyses were performed using SPSS, version 24 (IBM Corp. 2016), and JAMOVI computer software, version 1.6.3 (Jamovi Project 2019).

We invite readers to note that specific aspects of our experimental design were correlational and others causal. The manipulated variable was the participants' perspective in both scenarios, allowing causal inferences. However, the predicting effects of age, gender and empathy on participant's choices were correlational.

Results

Descriptive Statistics of Participants' Choices in Scenarios 1 and 2

Scenario 1. Overall, a significant difference was found in the participants' choices ($\chi^2 = 61.3$; $Df = 1$; $p < .001$): 74.5% made *group*-oriented choices (Cf. Figure 1). Men and women did not present significant differences in their choices in Scenario 1 ($\chi^2_{exact} = .59$; $Df = 1$; $p = .102$): 77% of women made *group*-oriented choices, and 67% of men made *group*-oriented choices (Cf. Figure 1). Furthermore, the attribution of penalties was divided into 72% of participants refusing to stop; 22% accepted taking the tardy student without conditions, and 6% accepted taking the tardy student with conditions (no difference was found between men and women: $\chi^2 = 3.36$; $Df = 2$; $p = .186$, Cf. Figure 2). ***Scenario 2***. In general, no significant difference was found in the participants' decisions ($\chi^2 = 1.73$; $Df = 1$; $p = .188$): 54% made *group*-oriented-choices and 46% made *self*-oriented choices (Cf. Figure 1). Significant gender differences were found ($\chi^2_{exact} = .531$; $Df = 1$; $p = .032$; $\phi = .139$): 58% of women made *group*-oriented choices, while men made 42% of *group*-oriented choices (Cf. Figure 1). Compared to Scenario 1, the attribution of penalties increased drastically: 53% still refused to stop the tramway; 29% stopped the tramway for them, without penalties; and 18% stopped the tramway for them, with penalties (Cf. Figure 2). No significant difference was found in penalty allotments between men and women ($\chi^2 = 4.57$; $Df = 2$; $p = .102$).

Participants showed choice stability between Scenarios 1 and 2 ($\chi^2_{exact} = 2.79$; $Df = 1$; $p < .001$; $\phi = .22$), as well for penalty conditions ($\chi^2 = 27.2$; $Df = 4$; $p < .001$; $\phi = .23$). Students' departments did not significantly influence choices in Scenarios 1 and 2 (respectively, $\chi^2 = 6.39$; $Df = 5$; $p = .270$; $\chi^2 = 2.43$; $Df = 5$; $p = .788$). ϕ coefficients between .20 and .40 represent moderate associations (Rea and Parker 1992), suggesting that choices were consistent in both scenarios. These relations also suggest that both scenarios present an acceptable reliability of dilemma-responses.

Gender, Empathy, and Choices in Scenarios 1 and 2

The Generalized Mixed Model analysis (LogLikel. = −307.4; $R^2_{Marg.} = .121$; $R^2_{Cond.} = .311$; AIC = 640.8; BIC; 695.9; Deviance = 475.9; $Df_{res.} = 497$) showed that the variable 'Scenario' was a significant predictor of the dependent variable 'STOP' ($B = 0.965$; $SE = 0.234$; $Exp(B) = 2.63$; $p < .001$): here, the probability to stop the tramway is 2.6 higher in Scenario 2 (see Table 2) compared to Scenario 1. Gender was also a significant predictor of the variable 'STOP' ($B = -0.809$; $SE = 0.312$; $Exp(B) = 0.445$; $p = .009$), meaning that the probability to stop the tramway was divided by 2 when the participant was a woman (see Table 2). More importantly, we found a significant three-way interaction effect between Scenario ✻ Gender ✻ AE ($B = 0.190$; $SE = 0.081$; $Exp(B) = 1.21$; $p = .019$; see Table 2 & Fig. 3a & b) to predict the variable 'STOP'. This three-way interaction revealed that: (1) in Scenario 1, men's AE linearly predicted *individual* and *group*-oriented choices. There was a probability of 20% to stop the tramway for men with −1 SD of AE levels (see Table B in Supplementary Materials), and this likelihood increased according to their AE levels (it reaches 55% for men with +1 SD of AE, and this continues to increase at +2 SD levels). Women's AE presented a flat predicting effect in stopping or not the tramway ($p = .234$, see Table C in Supplementary Materials). Independently of their AE levels, the probability of stopping the tramway was low and ranged from 16 to 24% (see Table B in Supplementary Materials). (2) In Scenario 2, the probability of stopping the tramway differed in men and women with lower (-1SD) and average AE scores (see Table D for gender differences in Supplementary Materials). At -1SD AE levels, while men had a higher

[7] Note that the model proceeds to an effect coding on the dummy variables. It recoded the gender and scenario variables into −0.5 and 0.5.

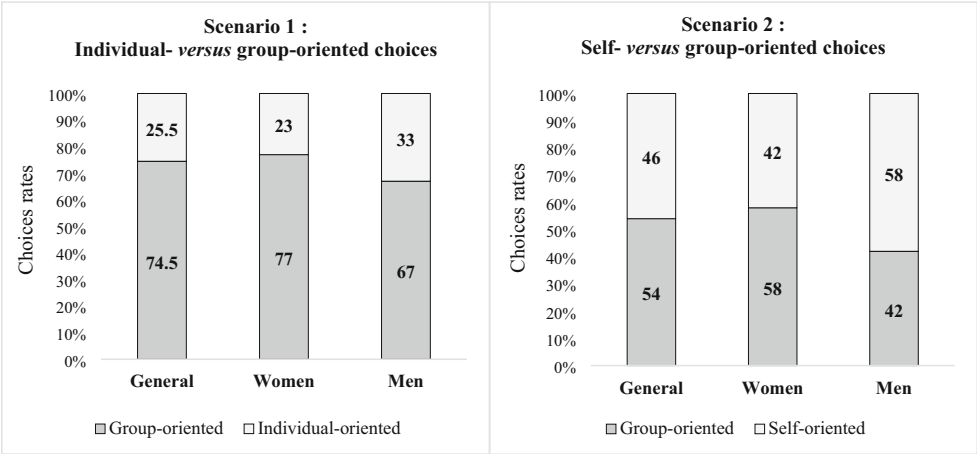


Fig. 1 Choice rates in scenarios 1 and 2. *Note.* This figure displays participants' choice rates for scenarios 1 and 2

propensity (around 60%) to stop the tramway (they were more likely to make *self*-oriented choices), women had a lower propensity (below 40%) of stopping it (they were more likely to make *group*-oriented choices).

Lastly, additional analyses showed that empathy did not predict penalty allotment in both scenarios (see Table E in Supplementary Materials). However, we found a marginal three-way interaction effect of Scenario ✻ Gender ✻ AE ($B = 0.189$; $SE = 0.10$; $Exp(B) = 1.21$; $p = .057$; see Table F in Supplementary Materials); showing that, in Scenario 1, higher men's AE levels increased the probability of stopping the tramway without imposing any penalties. A simple effect analysis showed that the linear regression slope of men's AE in Scenario 1 was significant ($p = .043$; see Table Fa & Figures in Supplementary Materials).

Note. This figure displays participants' penalty rates for scenarios 1 and 2.

Fig. 2 Choice rates of penalties in scenarios 1 and 2. *Note.* This figure displays participants' rates of penalties for scenarios 1 and 2

Discussion

The *Individual*, *Self*, or *Group*-Oriented Choices in Trolley-Like Problems

As found by other authors (Di Nucci 2013; Swann et al. 2010), we have investigated whether people make more *group*-oriented (*utilitarian*) choices than *individual* or *self*-oriented (*non-utilitarian*) choices in social dilemmas. According to the philosophers' theory of ethics (e.g., Jeremy Betham, John Stuart Mill, and Henri Sidgwick), *utilitarianism* is equivalent to preserve a group entity more frequently than a single isolated person because, following a cost-benefit analysis, the greatest overall utility is to sacrifice one person rather than a group to maximize the aggregated happiness (also called "the Greatest Happiness Principle"). The findings confirmed Hypothesis 1a: most participants (75%) made more *group*-oriented choices in the *no-direct-impact* scenario (Scenario 1). However, in the *direct-impact* scenario (Scenario 2), we did not find this pattern. Indeed, whenever participants were the tardy student, the frequency of making the tramway stop dropped down to a fifty-fifty probability. This difference is caused by the participant's degree of implication in the scenarios and, more precisely, to the participant's perspective in these scenarios. As Greene et al. (2001) mentioned, *impersonal* scenarios (which can be compared to Scenario 1) typically involve less social-emotional responses and more cognitive processes than *personal* scenarios. Following our results and previous findings, it is clear that the higher degree of *group*-oriented choices depends on the participant's degree of implication (as predicted by the Dual-Process Model (Greene et al. 2001)) and on the extent of consequences s/he will experience. In that sense, despite the higher rate of *group*-oriented (*utilitarian*) choices in Scenario 1, this percentage is lower than previous findings (e.g., Navarrete et al. (2012) found around 90% *utilitarian* choices); suggesting that causing inconvenience to others leads to lower degrees of *group*-oriented decisions than killing people. In contrast, in Scenario

Table 2 Coefficients of the general mixed model analysis

			95% CI			95% Exp(B) CI			
Names	**E**	**SE**	**Lower**	**Upper**	**Exp(B)**	**Lower**	**Upper**	**z**	**p**
Intercept	−0.519	0.157	−0.826	−0.211	0.595	0.438	0.810	−3.30	<.001
Scenario	0.965	0.234	0.508	14.2	2.63	1.66	4.15	4.13	<.001
Gender	−0.809	0.312	−14.2	−0.198	0.445	0.242	0.820	−2.60	0.009
Age	0.020	0.043	−0.063	0.104	1.02	0.939	1.11	0.474	0.636
AE	0.021	0.024	−0.027	0.068	1.02	0.974	1.07	0.852	0.394
CE	−0.050	0.036	−0.120	0.020	0.951	0.887	1.02	−1.39	0.164
Scenario ✻ AE	−0.057	0.048	−0.151	0.037	0.945	0.860	1.04	−1.18	0.237
Gender ✻ AE	−0.037	0.039	−0.114	0.039	0.963	0.892	1.04	−0.954	0.340
Scenario ✻ CE	−0.044	0.072	−0.184	0.097	0.957	0.832	1.10	−0.609	0.542
Gender ✻ CE	0.036	0.062	−0.085	0.157	1.04	0.919	1.17	0.588	0.557
Scenario ✻ Gender ✻ AE	0.190	0.081	0.031	0.349	1.21	1.03	1.42	2.35	0.019
Scenario ✻ Gender ✻ CE	−0.044	0.124	−0.286	0.199	0.957	0.751	1.22	−0.352	0.725

Note. This table displays all predictors of the general mixed model analysis. Abbreviations: AE = Affective Empathy and CE = Cognitive Empathy

2, the participant personally and directly suffers from the consequences of his/her decision: attending the lecture or not. Thus, in this scenario, when the participant stops the tramway, s/he decides to pursue egocentric purposes.

As predicted by previous findings and theory, people make more *group*-oriented choices in the *no-direct-impact* scenario than the *direct-impact* scenario. Indeed, it elicits fewer emotions and involves more cognitive processes (as predicted by Greene et al.'s (2001) dual-process model). However, considering the *direct-impact* scenario, Crockett et al.'s (2014) showed that causing harm appears to be set on the paradigm that most people valued others' pain more than their own, leading them to be 'hyperaltruistic' and, thus, *utilitarianists* (in this study, participants had to choose between causing harm to themselves or others). From the causing inconvenience perspective, the bad outcomes aversion for others is evaluated differently because the cost-benefit analysis produces mitigated decisions. Thus, despite Kahneman's statement (Kahneman 2013) that people dislike being responsible for affecting others, overall, there is a fifty-fifty probability of making *self*- or *group*-oriented choices (even though gender differences were found).

Considering gender differences in decision-making, although we did not find differences in Scenario 1, men and women presented a reversed pattern of choices in Scenario 2 (which contradicts Hypothesis 1b). Women made higher *group*-oriented (*utilitarian*) choices than men when the consequences of their decisions directly impacted them. This result is the opposite of Fumagalli et al.'s findings (Fumagalli et al. 2010a; Fumagalli et al. 2010b), where males presented more *utilitarian* choices in *personal* moral dilemmas. To explain this difference, in our study, we did not ask participants to violate their moral standards by killing someone. Participants were asked to decide to stop or not the tramway, a choice that will cause inconvenience to others. Another explanation might be that participants changed their perspective

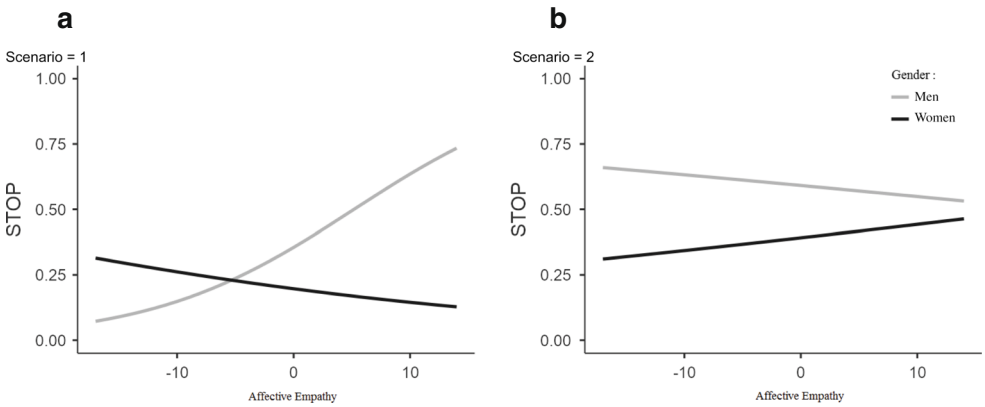


Fig. 3 a & b. Interaction between gender and affective empathy in scenarios 1 and 2. *Note*. These figures display the predicting effect of affective empathy on the variable 'STOP' for men and women in scenarios 1 and 2

Note. These figures display the predicting effect of affective empathy on the variable 'STOP' for men and women in scenarios 1 and 2.

by being themselves the tardy student in Scenario 2. In this kind of scenario, men made more egocentric and women more altruistic choices. Indeed, women might be more sensitive than men to the inconvenience they caused and will be more inclined to care for the group over themselves. This interpretation suggests that men and women might use different coping strategies to reduce the cognitive dissonance associated with their choices. However, we have to keep in mind that the slight gender difference observed in our sample aligns with the arguments made by Lapsley (1996) and Walker (1991) that the degree to which gender (per se) affects moral judgment is small.

The Frequencies of Penalty Allotment in Trolley-Like Problems

Interestingly, when participants were asked to choose whether or not to impose penalties on the tardy student (Scenario 1), the frequency of penalty allotment was low (6%). However, this frequency tripled when participants were themselves the late student (Scenario 2), confirming Hypothesis 2a. The low frequency of penalty allotment in Scenario 1[8] joins the Trémolière and Djeriouat's statement (2016) that individuals assigned much less guilt and were much less severe to punish non-causally responsible people, even if that person intended harming someone. However, our results indicate that people are much more inclined to compensate others when they inconvenience them than they expect from someone causing inconvenience to compensate others.

Empathy did not significantly predict penalty allotment in both scenarios, which contradicts Hypothesis 2b. Instead, men's affective empathy marginally predicted the probability of stopping the tramway without imposing any penalties. In our opinion, this effect stresses that empathy is a prosocial feeling (Davis 2015), promoting people to help each other (even if it seems especially true for men immersed in a *no-direct-impact* scenario). However, empathy does not explain the increase in penalty allotment in Scenario 2. In our opinion, a great candidate to explain this increase would be emotion, like guilt. As Choe and Min (2011) showed, several emotions are elicited during judgment: anger, sadness, disgust, anxiety, empathy, and especially guilt. In that sense, other authors compared the role of guilt and shame (two emotions considered as "moral emotions"; see Tangney et al. 2007; Haidt 2003) on reparative behaviors (Ghorbani et al. 2013). They found that guilt produced stronger levels of compensation than shame. Furthermore, "*the closer people feel toward the target, the more likely they will feel ashamed and guilty about the past harmful behavior, and therefore offer higher compensation"* (Ghorbani et al. 2013, p.9).

Indeed, emotions can foster the motivation to "retribute" others for the inconvenience caused, and according to the available possibilities of the situation, this person will compensate others by producing prosocial behaviors (i.e., *active* production) and/or by taking penalties (i.e., *passive* production). The underlying motives are mainly egoistic but indirectly altruistic. As Batson et al. (2015 p.3) stated, "*altruism and egoism refer to a motivational state […]; and, for each, the ultimate goal is to increase someone's welfare*". These strategies (i.e., *active* and/or *passive* production of compensations) might lead people to maintain a positive self-image when behaving selfishly (Mazar et al. 2008; Benabou and Tirole 2011).

The Specific Role of Empathy, Gender, and Scenario in Trolley-Like Dilemmas

This research's main purpose was to determine whether empathy leads participants to make specific decisions in everyday Trolley-like Problems. Following Decety and Cowell's (2014) arguments, previous findings (Gleichgerrcht and Young 2013; Patil and Silani 2014), and predictions from the Dual-Process Model (Greene et al. 2001, see Table A in Supplementary Materials), we expected to find a significant predictive effect of affective empathy on *utilitarian* moral decisions (Hypothesis 3). Our results partly replicated those findings and showed different predicting effects of affective empathy in both scenarios according to gender. We found that men's affective empathy levels (i.e., men's ability to feel what others feel) predicted *individual* and *group*-oriented choices in the *no-direct-impact* scenario. Indeed, the probability of stopping the tramway (i.e., making an *individual*-oriented choice) increased as a function of men's affective empathy. Furthermore, at lower and average affective empathy levels, we found different predicting effects of men and women's affective empathy on choices in Scenario 2. At those affective empathy levels, men were more likely to make *self*-oriented decisions, and women were more likely to make *group*-oriented decisions.

The affective empathy relation with utilitarianism is very well established in literature but rarely studied in a three-way interaction model, including the gender, scenarios, and empathy variables. As mentioned above, previous research showed that people with lower affective empathy levels (e.g., autism or psychopathy) display more utilitarian choices (Decety and Cowell 2014; Gleichgerrcht and Young 2013; Patil and Silani 2014). However, this three-way interaction effect raised several interpretations.

First, because the participant and the target shared several features (i.e., age and gender), we believe that the sensitivity to similarities might mediate the relation between men's

[8] In both scenarios, the target or the participant has no intention to harm or cause inconvenience to others. For Shultz and Wells (1985), intention is a critical component in morality.

affective empathy and choices in *no-direct-impact* situations. Preston and de Waal (2002) described *similarity* (i.e., the perceived overlap between subject and object, such as personality, age, gender) as one of the fundamental factors of empathy with *familiarity*, *learning*, *past experience*, and *salience*. Moreover, Eagly and Crowley (1986) showed that men are more homophilic than women in helping behaviors, even though this homophilic effect has been controverted (see George et al. 1998). In Trolley Problems, Skulmowski et al. (2014) showed that men were more homophilic in sacrificing a male target than women. Another study showed that the feeling of relatedness was a significant mediating variable between antagonistic personality and moral outrage perception (Hart et al. 2020), suggesting that personality traits and moral perception can be mediate by similarity. At this stage of understanding, further investigation is required to clarify this presumed effect of similarity on men's affective empathy and decision-making (see Table F in Supplementary Materials).

A second interpretation is that the attentional focus might be driven differently by men's affective empathy, leading them to make different choices. Wu et al. (2017) provided strong evidence that state empathy guides and regulates attention processes. However, currently, the authors had only studied the link between attention and choices. Kastner (2010, p. 123) showed that "*people avoid looking at information that is inconsistent with their intuitively formed moral judgment.*" In this study, more visual attention was devoted to the saved than the killed (the author examined fixation behavior after the participant's decision). In contrast, Skulmowski et al. (2014) demonstrated that participants in virtual-reality Trolley Problems spend more time looking at the killed, suggesting that they would like to reassure themselves of making the right decision (the authors measured participant's gaze during the decision process). In our scenarios, participants had no visual cues and presumably created a mental representation of the social dilemmas. Therefore, instead of giving a specific answer, these results raise an important question: is there a causal role of affective empathy levels on attentional processes (creating an attentional bias toward the victim(s) or the saved) that influences the decision-making process in Trolley (and Trolley-like) Problems? For example, in a *no-direct-impact* context (when a target is a man), men's attentional focus might switch from a group to one individual depending on their affective empathy levels. As Slovic (2007, p. 90) explained, "*the emotion or affective feeling* [toward a victim] *is greatest at N = 1 but begins to decline at N = 2 and collapses at some higher value of N that becomes simply "a statistic*"; and it seems particularly true for men with higher affective empathy. In addition, Lickel et al. (2000) have argued that groups are more abstract than individuals and generally elicit less empathy than individuals (Cameron and Payne 2011). Following Slovic' and other authors' explanations, we suggest that the group is considered as an undifferentiated mass for men with higher affective empathy, leading them to devote more attention to the individual and make more *individual*-oriented decisions.

A third interpretation draws on Takamatsu's findings (Takamatsu 2018). She demonstrates that, regardless of gender, individuals with lower empathic concern are more likely to make utilitarian choices because they feel lower empathy for the victim and higher empathy for the saved. We can make a parallel between these results and our findings. Men in Scenario 1 and women in Scenario 2 had probably higher empathy for the saved and lower empathy for the victim, which drive their decisions toward the group. However, this interpretation does not explain why men with lower affective empathy levels made more *self*-oriented choices in Scenario 2. Indeed, contrary to Scenario 1, at lower and average affective empathy levels, we found opposite predicting effects of men and women's affective empathy on choices in Scenario 2. At those affective empathy levels, men were more likely to make *self*-oriented choices (i.e., corresponding to a selfish decision), and women were more likely to make *group*-oriented choices (i.e., corresponding to an *altruistic* decision). These results obtained in Scenario 2 are concordant with previous findings showing that women are more socially-oriented while men are more selfish (Eckel and Grossman 1998; Brañas-Garza et al. 2018).

Lastly, the Dual-Process Model of moral judgment (Greene et al. 2001) predicts that scenarios eliciting lower emotional and higher cognitive processes (like *impersonal* scenarios or Scenario 1) produce higher rates of *group*-oriented decisions. On the other hand, scenarios eliciting higher emotional and lower cognitive processes (like *personal* scenarios or Scenario 2) produce lower rates of *group*-oriented decisions. By measuring affective empathy, the model can predict four different patterns of answers in these two types of scenarios (see Table A in Supplementary Materials). Depending on gender, our findings fit with these predictions, except for prediction 2, where men and women higher in affective empathy had a fifty-fifty probability of making *group*-oriented choices. Prediction 1 was correct for men but not for women, and prediction 3 was correct for women but not for men. Overall, even though the specific role of affective empathy in gender requires further investigation, these results provide evidence that Greene et al.'s (2001) model is efficient in predicting individuals' moral decision-making.

Conclusion

In conclusion, the development of everyday Trolley-like Problems (increasing ecological validity) yields significant benefits. Widely used, they will bring more consistent and robust evidence of individuals' moral decision-making in everyday life situations and the significant influence of constructs, such as empathy, on this decision-making process.

The novelties brought by our investigation were: (1) to put the participant into different perspectives of everyday Trolley-like Problems (which significantly decreased *group*-oriented choices); (2) to assess the association between empathy and moral decision-making by using the BES; and (3) to explore the frequency of penalty allotment associated with the participant's moral choices, and the predicting effect of affective empathy. Although some of our results were concordant with previous findings, our Trolley Problems variants showed some discrepancies. We demonstrated that moving from a perspective to another in moral dilemmas (from Scenario 1 to Scenario 2, or conversely) caused substantial changes in participants' decision-making. We also observed that, despite the higher rates of *group*-oriented choices in Scenario 1 (75%), this percentage is inferior to Navarette et al. 's (Navarrete et al. 2012) percentage (around 90%). Furthermore, contrary to previous findings, women from our sample presented more *group*-oriented choices in the *direct-impact* scenario (Scenario 2) than men. Also, by using different tools (i.e., Trolley-like Problems and the BES), we showed that the probability of stopping the tramway increased as a function of men's affective empathy. Lastly, we found that people seem to be less inclined to impose penalties on someone who causes inconvenience to others than they are for themselves. Overall, these findings are in line with current attempts to make Trolley Problems more familiar to everyday life situations. The variants proposed in this study raise new questions and allow researchers to explore new approaches (e.g., the role played of similarity and attention, or penalty allotments), giving new research horizons. Furthermore, these results demonstrated that investigating the role of empathy is still a challenge for understanding and extending our knowledge on moral decision-making. Lastly, to develop a cataphatic theoretical framework of moral psychology (or psychology of morality), it seems necessary to adopt a new decision-making nomenclature (i.e., *individual*, *self*, or *group*-oriented choices).

Limitations and Direction for Future Research

This study presents several limitations. First, like most experimental designs using Trolley Problems and other variants, the findings obtained from this correlational design do not allow to make causal inferences. Indeed, we cannot consider that affective empathy causes *individual* or *group*-oriented choices. Second, our study is focused on university students. We reasonably assume that our sample is representative of the general population, but further investigations can be extended to the general population. Third, we did not assess moral decisions in a cross-gender way: in our study, women participants received women's scenarios, and men received men's scenarios.

We invite readers to keep in mind that the three-way interactive effect of AE, gender, and scenario on moral decision-making is to put in perspective with the other regressors present in the equation (notably, cognitive empathy), even if they were not significant. Indeed, this interaction is significant when all other parameters are equal to their means. Furthermore, following Greene et al.'s findings (Greene et al. 2008), cognitive processes have a crucial impact on utilitarian choices. Therefore, it remains surprising that cognitive empathy does not significantly affect moral decision-making in Trolley Problems and its variants, suggesting that theoretical gaps are still to be explored.

These findings shed light on the importance of using everyday Trolley-like Problems. It might provide ecological evidence of moral dysfunction when investigating psychopathological populations, like psychopathy or the Dark Tetrad. We also proposed some interpretations where the presumed mechanisms were not tested in our study. Therefore, it would be important for future research to investigate the following leads: (1) the potential role of emotions in predicting penalty allotment, (2) the mediating effect of similarity for men on moral decision-making, (3) the causal influence of affective empathy on attention, and (4) exploring some theoretical gaps to explain the absence of influence of cognitive empathy in participants' choices. We believe that everyday Trolley-like Problems will provide more concrete and generalizable evidence of constructs influencing decision-making in social dilemmas.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12144-021-01566-1) contains supplementary material, which is available to authorized users.

Open Practices Data are available online: DOI: https://doi.org/10.17605/OSF.IO/5Q6XN.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

Armstrong, J., Friesdorf, R., & Conway, P. (2019). Clarifying gender differences in moral dilemma judgments: The complementary roles of harm aversion and action aversion. *Social Psychological and Personality Science, 10*(3), 353–363. https://doi.org/10.1177/1948550618755873.

Balash, J., & Falkenbach, D. M. (2018). The ends justify the meanness: An investigation of psychopathic traits and utilitarian moral endorsement. *Personality and Individual Differences, 127*, 127–132. https://doi.org/10.1016/j.paid.2018.02.009.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*(1), 154–161. https://doi.org/10.1016/j.cognition.2011.05.010.

Batson, C. D., Lishner, D. A., Stocks, E., & L. (2015). The empathy-altruism hypothesis. In A. Schroeder & W. G. Graziano (Eds.), *The Oxford handbook of prosocial behavior*. Oxford: Oxford library of psychology. https://doi.org/10.1093/oxfordhb/9780195399813.013.023.

Batson, C. D., Sager, K., Garst, E., Kang, M., Rubchinsky, K., & Dawson, K. (1997). Is empathy-induced helping due to self-other merging? *Journal of Personality and Social Psychology, 73*, 495–509.

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass, 8*(9), 536–554. https://doi.org/10.1111/spc3.12131.

Benabou, R., & Tirole, J. (2011). Identity, morals and taboos: Beliefs as assets. *Quarterly Journal of Economics, 126*(2), 805–855. https://doi.org/10.1093/qje/qjr002.

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science, 29*(7), 1084–1093. https://doi.org/10.1177/0956797617752640.

Brañas-Garza, P., Capraro, V., & Rascon-Ramirez, E. (2018). Gender differences in altruism on mechanical Turk: Expectations and actual behaviour. *Economics Letters, 170*, 19–23. https://doi.org/10.1016/j.econlet.2018.05.022.

Bruers, S., & Braeckman, J. (2014). A review and systematization of the trolley problem. *Philosophia, 42*(2), 251–269. https://doi.org/10.1007/s11406-013-9507-5.

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development, 53*(2), 413–425. https://doi.org/10.2307/1128984.

Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology, 100*, 1–15.

Carré, A., Stefaniak, N., D'Ambrosio, F., Bensalah, L., & Besche-Richard, C. (2013). The basic empathy scale in adults (BES-A): Factor structure of a revised form. *Psychological Assessment, 25*(3), 679-691. DOI: 10.1037:a0032297.

Choe, S. Y., & Min, K. H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment & Decision Making, 6*(7), 580–592.

Copraro, V., & Sippel, J. (2017). Gender differences in moral judgment and the evaluation of gender-specified moral agents. *Cognitive Processing, 18*, 399–405. https://doi.org/10.1007/s10339-017-0822-9.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences, 111*(48), 17320–17325. https://doi.org/10.1073/pnas.1408988111.

D'Ambrosio, F., Olivier, M., Didon, D., & Besche, C. (2009). The basic empathy scale: A French validation of a measure of empathy in youth. *Personality and Individual Differences, 46*(2), 160–165. https://doi.org/10.1016/j.paid.2008.09.020.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126.

Davis, M. H. (2015). Empathy and prosocial behavior. In A. Schroeder & W. G. Graziano (Eds.), *The Oxford handbook of prosocial behavior*. Oxford: Oxford library of psychology. https://doi.org/10.1093/oxfordhb/9780195399813.013.026.

de Waal, F. B. M., & Suchak, M. (2010). Prosocial primates: Selfish and unselfish motivations. *Philosophical Transactions of the Royal Society-B, 365*, 2711–2722.

Decety, J. (2015). The neural pathways, development and functions of empathy. *Current Opinion in Behavioral Sciences, 3*, 1–6.

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science, 9*(5), 525–537. https://doi.org/10.1177/1745691614545130.

Di Nucci, E. (2013). Self-sacrifice and the trolley problem. *Philosophical Psychology, 26*(5), 662–672. https://doi.org/10.1080/09515089.2012.674664.

Djeriouat, H., & Trémolière, B. (2014). The dark triad of personality and utilitarian moral judgment: The mediating role of honesty/humility and harm/care. *Personality and Individual Differences, 67*, 11–16. https://doi.org/10.1016/j.paid.2013.12.026.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100*(3), 283–308.

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal, 108*, 726–735.

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition, 123*(3), 434–441. https://doi.org/10.1016/j.cognition.2012.02.001.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. In Foot (1977). Virtues and vices and other essays in moral philosophy. Oxford: Basil Blackwell, pp.19–32.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality and Social Psychology Bulletin, 41*(5), 696–713. https://doi.org/10.1177/0146167215575731.

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., Lucchiari, C., Consonni, D., Nordio, F., Pravettoni, G., & Cappa, S. (2010a). Gender-related differences in moral judgments. *Cognitive Processing, 11*(3), 219–226. https://doi.org/10.1007/s10339-009-0335-2.

Fumagalli, M., Vergari, M., Pasqualetti, P., Marceglia, S., Mameli, F., Ferrucci, R., Mrakic-Sposta, S., Zago, S., Sartori, G., Pravettoni, G., Barbieri, S., Cappa, S., & Priori, A. (2010b). Brain switches utilitarian behavior: Does gender make the difference? *PLoS One, 5*(1), e8865. https://doi.org/10.1371/journal.pone.0008865.

Gallucci, M. (2019). GAMLj: General analyses for linear models. [Jamovi module]. Retrieved from https://gamlj.github.io/.

Gao, Y., & Tang, S. (2013). Psychopathic personality and utilitarian moral judgment in college students. *Journal of Criminal Justice, 41*(5), 342–349. https://doi.org/10.1016/j.jcrimjus.2013.06.012.

George, D., Carroll, P., Kersnick, R., & Calderon, K. (1998). Gender-related patterns of helping among friends. *Psychology of Women Quarterly, 22*(1), 685–704. https://doi.org/10.1111/j.1471-6402.1998.tb00185.x.

Ghorbani, M., Liao, Y., Çayköylü, S., & Chand, M. (2013). Guilt, shame, and reparative behavior: The effect of psychological proximity. *Journal of Business Ethics, 114*(2), 311–323. https://doi.org/10.1007/s10551-012-1350-2.

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS One, 8*(4), e60418. https://doi.org/10.1371/journal.pone.0060418.

Gleichgerrcht, E., Torralva, T., Rattazzi, A., Marenco, V., Roca, M., & Manes, F. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *Social Cognitive and Affective Neuroscience, 8*(7), 780–788. https://doi.org/10.1093/scan/nss067.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11*(8), 322–323.

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.

Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517–523. https://doi.org/10.1016/s1364-6613(02)02011-9.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford: Oxford University Press.

Hart, W., Kinrade, C., Tortoriello, G. K., Richardson, K., & Breeden, C. J. (2020). Identifying with the "bad" guy: A novel account for apparent moral-judgment deficiencies in antagonistic personality. *Personality and Individual Differences, 166*(1), 110161. https://doi.org/10.1016/j.paid.2020.110161.

Hoffman, M. L. (2000). Empathy and moral development: Implications for caring and justice. Cambridge University Press.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33*, 307–316.

IBM Corp. Released. (2016). *IBM SPSS statistics for windows, version 24.0*. Armonk: IBM Corp.

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence, 29*, 589-611. DOI: 10.1016:J.adolescence.2005.08.010.

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience, 10*(5), 551–560. https://doi.org/10.1080/17470919.2015.1023400.

Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193–209. https://doi.org/10.1016/j.cognition.2014.10.005.

Kahneman, D. (2013). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kastner, R. M. (2010). Moral judgments and visual attention: An eye-tracking investigation. *Chrestomathy: Annual Review of Undergraduate Research, School of Humanities and Social Sciences, School of Languages, Cultures, and World Affair, 9*, 114–128.

Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience, 7*(6), 708–714. https://doi.org/10.1093/scan/nsr048.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature, 446*(7138), 908–911. https://doi.org/10.1038/nature05631.

Lapsley, D. K. (1996). *Moral psychology*. Boulder: Westview Press.

Li, Y., & Liu, A.-S. (2017). Moral judgment and its relationship with theory of mind in patients with autism spectrum disorder. *Chinese Mental Health Journal, 31*(5), 366–370.

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology, 78*, 223–246.

Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making, 27*, 57–65. https://doi.org/10.1002/bdm.1782.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*, 633–634.

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem". *Emotion, 12*(2), 364–370. https://doi.org/10.1037/a0025561.

Patil, I. (2015). Trait psychopathy and utilitarian moral judgment: The mediating role of action aversion. *Journal of Cognitive Psychology, 27*(3), 349–366. https://doi.org/10.1080/20445911.2015.1004334.

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology, 5*, 501. https://doi.org/10.3389/fpsyg.2014.00501.

Pletti, C., Lotto, L., Buodo, G., & Sarlo, M. (2017). It's immoral, but I'd do it! Psychopathy traits affect decision-making in sacrificial dilemmas and in everyday moral situations. *British Journal of Psychology, 108*(2), 351–368. https://doi.org/10.1111/bjop.12205.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1–72. https://doi.org/10.1017/S0140525X02000018.

Ramsøy, T. Z., Skov, M., Macoveanu, J., Siebner, H. R., & Fosgaard, T. R. (2015). Empathy as a neuropsychological heuristic in social decision-making. *Social Neuroscience, 10*(2), 179–191. https://doi.org/10.1080/17470919.2014.965341.

Raosoft. (2004). Raosoft Sample Size Calculator. Raosoft, Inc., Seattle. http://www.raosoft.com/samplesize.html

Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research*. San Francisco: Jossey–Bass.

Shultz, T. R., & Wells, D. (1985). Judging the intentionality of action-outcomes. *Developmental Psychology, 21*(1), 83–89.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye-tracking study. *Frontiers in Behavioral Neurosciences, 8*(426), 1–16. https://doi.org/10.3389/fnbeh.2014.00426.

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision making, 2*, 79–95.

Swann, W. B., Gómez, A., Dovidio, J. F., Hart, S., & Jetten, J. (2010). Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science, 21*(8), 1176–1183. https://doi.org/10.1177/0956797610376656.

Takamatsu, R. (2018). Turning off the empathy switch: Lower empathic concern for the victim leads to utilitarian choices of action. *PLoS One, 13*(9), e0203826. https://doi.org/10.1371/journal.pone.0203826.

Takamatsu, R. (2019). Personality correlates and utilitarian judgments in the everyday context: Psychopathic traits and differential effects of empathy, social dominance orientation, and dehumanization beliefs. *Personality and Individual Differences, 146*, 1–8. https://doi.org/10.1016/j.paid.2019.03.029.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual. Review of Psychology, 58*, 345–372. https://doi.org/10.1146/annurev.psych.56.091103.070145.

The Jamovi Project (2019). *Jamovi, version 1.6.3* [Computer Software]. Retrieved from https://www.jamovi.org

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59*, 204–217.

Trémolière, B., & Djeriouat, H. (2016). The sadistic trait predicts minimization of intention and causal responsibility in moral judgment. *Cognition, 146*, 158–171. https://doi.org/10.1016/j.cognition.2015.09.014.

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment, 23*(2), 135–149.

Vyas, K., Jameel, L., Bellesi, G., Crawford, S., & Channon, S. (2017). Derailing the trolley: Everyday utilitarian judgments in groups high versus low in psychopathic traits or autistic traits. *Psychiatry Research, 250*, 84–91. https://doi.org/10.1016/j.psychres.2017.01.054.

Walker, L. J. (1991). Sex differences in moral reasoning. In W.M., Kurtines & J. L., Gewirtz (Eds.), Handbook of moral behavior and development (pp.333–364). Hillsdale, NJ: Erlbaum.

Wu, L., Kirmse, U., Flaisch, T., Boiandina, G., Kenter, A., & Schupp, H. T. (2017). Empathy, pain and attention: Cues that predict pain stimulation to the partner and the self capture visual attention. *Frontiers in Human Neuroscience, 11*, 465. https://doi.org/10.3389/fnhum.2017.00465.

Synthèse de l'étude 1 :

Les résultats de cette première étude démontrent une innovation majeure dans le domaine : un effet conséquent de la perspective du participant dans les scénarios sur la prise de décision. En effet, les choix utilitaires passent de ≈75% dans la situation impersonnelle (i.e., quand le(a) participant(e) incarne le(a) conducteur(trice)) à 54% dans la situation personnelle (i.e., quand le(a) participant(e) incarne le(a) retardataire).

Quant à l'empathie affective, on observe un effet prédicteur (assez faible) uniquement à travers une interaction triple entre le scénario, le genre et l'empathie affective. Cet effet montre un effet de régression linéaire de l'empathie affective sur les choix utilitaires chez les hommes (lorsqu'ils incarnent le conducteur de tram). Avec une faible empathie affective, les participants vont avoir tendance à moins s'arrêter et, à mesure que leur empathie affective augmente, ils vont avoir tendance à s'arrêter davantage pour charger le retardataire. Toutefois, il n'y a aucun effet de l'empathie affective pour les femmes dans ce scénario. Enfin, dans le scénario où les participant(e)s incarnent le(a) retardataire, c'est uniquement à de faibles niveaux d'empathie affective que l'on observe un effet significatif sur les choix utilitaires. Effectivement, la probabilité de faire arrêter le tram était plus importante chez les hommes que chez les femmes (ne pas faire arrêter le tram correspond à un choix utilitaire). Ce dispositif expérimental permet de montrer que les prédictions faites à partir du DPM pour l'empathie affective se sont avérées partiellement correctes, en ce sens où elles vont dépendre du genre.

Toutefois, ces résultats soulèvent deux problèmes majeurs : le premier étant un effet particulièrement faible de l'empathie affective pour prédire les choix utilitaires et le second étant l'absence d'effet prédicteur significatif de l'empathie cognitive sur les choix. De fait, comme le mentionne le DPM, ce sont les processus émotionnels et cognitifs qui jouent un rôle majeur dans les jugements et prise de décision morales. Or, comme soulevé dans notre article, pour quelle raison l'empathie cognitive (i.e., le fait de se mettre à la place d'autrui pour comprendre ses émotions) n'intervient pas dans la prise de décision ? Cette question et ces points fondamentaux seront abordés dans notre méta-analyse et revue systématique.

Chapitre 3 : Deuxième étude

Etude 2 : étude des liens entre empathie, perspective du participant, traits borderline et choix moraux dans des dilemmes de type trolley

L'étude précédente démontrait un effet majeur de la perspective du participant dans les scénarios sur les choix utilitaires. De plus, un effet, mineur quant à lui, de l'empathie affective avait été observé, qui variait selon le genre et la perspective du participant.

Sur cette base, nous avons, dans un premier temps, construit un nouveau matériel de type trolley de la vie quotidienne (baptisé « le dilemme de l'entreprise »). Dans un second temps, nous avons élargi la perspective que pouvait prendre le(a) participant(e) dans les scénarios à une troisième : celui(celle) parmi les cinq. Dès lors, le(a) participant(e) incarnait à présent trois perspectives dans les scénarios. Ensuite, nous avons utilisé un dilemme dit « classique », le dilemme du chirurgien (*The transplant dilemma*), présentant la même structure que le dilemme de l'entreprise (i.e., trois perspectives). La différence était tout de même majeure car, dans le dilemme classique, le(a) participant(e) devait décider de sacrifier la vie d'un individu pour en sauver cinq (ou ne rien faire) ; c'est ce que l'on a nommé un *causing harm scenario*. En contraste, dans le dilemme de l'entreprise, l'individu devait causer un désagrément majeur à un individu (le licencier) pour sauver la situation professionnelle de cinq autres travailleurs ; c'est ce que l'on a nommé un *causing inconvenience scenario*. Enfin, nous avons investigué l'effet prédicteur des traits de personnalité borderline et de l'empathie dans la prise de décision.

Comme mentionné dans l'introduction, la psychopathie est apparue comme l'une des premières psychopathologies qui a été investiguée dans la prise de décision morale. Un des symptômes cardinaux de cette psychopathologie est l'émoussement affectif. Ainsi, les auteurs ayant investigué cette population soutenaient que le fait de ressentir peu d'émotions devrait conduire les personnes psychopathes à réaliser davantage de choix utilitaires (comme préconisé par le DPM). Parallèlement, l'empathie affective des personnes psychopathes apparaît particulièrement faible - comme soutenu par une récente méta-analyse (Burghart & Mier, 2022) - ce qui les conduiraient à réaliser davantage de choix utilitaires. Notre point de réflexion s'est dès lors porté sur la population clinique présentant

des caractéristiques émotionnelles et empathiques opposées ; à savoir, les personnes présentant un trouble de personnalité borderline.

Le trouble de la personnalité borderline est caractérisé par un pattern persistant d'instabilité au niveau des affects, des relations interpersonnelles, de l'image personnelle et on constate chez ces personnes une impulsivité marquée. En comparaison avec la psychopathie, certains symptômes apparaissent similaires (par exemple, l'impulsivité), mais d'autres distinguent foncièrement ces deux psychopathologies : les psychopathes vont présenter une insensibilité émotionnelle (Hare, 1996 ; Hare & Neumann, 2009) et une hypo-réactivité émotionnelle en comparaison aux personnes borderline (Herpertz et al., 2001), tandis que les personnes borderline vont présenter une réactivité émotionnelle plus marquée (Crowell et al., 2009). D'ailleurs, certains auteurs soutiennent que l'intensité et la labilité émotionnelle et affective sont des symptômes cardinaux chez les personnes borderline (Richetin et al., 2017 ; Southward & Cheavens, 2018), et ce indépendamment de l'âge (Peckham et al., 2020).

Au niveau de l'empathie, deux récentes revues systématiques et méta-analyses montrent que les personnes borderline vont présenter de plus hauts niveaux en empathie affective et de plus bas niveaux en prise de perspective[5] (Salgado et al., 2020), tandis que les personnes psychopathes vont présenter de bas niveaux en empathie affective et en prise de perspective (Burghart & Mier, 2022). De ce fait, à niveau d'empathie cognitive égal, on constate deux populations polarisées au niveau de leur empathie affective (qui correspond à notre variable d'intérêt pour étudier les choix utilitaires).

Prolongeons à présent la logique du DPM : pour une population qui va présenter comme caractéristiques centrales une intensité et labilité émotionnelle importantes et de plus hauts niveaux d'empathie affective, nous devrions nous attendre à observer une relation négative entre les choix utilitaires et les traits de personnalité borderline (i.e., plus les traits sont présents, moins il y aura de propension à réaliser un choix orienté vers le groupe). Dans cette seconde étude, nous avons étudié les liens entre les choix orientés vers le groupe (choix utilitaire), l'empathie, la perspective incarnée par le(a) participant(e) et les traits de personnalité borderline au sein de la population générale. En effet, nous cherchions, dans un premier temps, à valider nos hypothèses sur une population générale très large (N > 400) pour pouvoir les tester, dans un second temps, sur une population

[5] Il s'agit d'une sous-dimension d'empathie cognitive.

clinique plus restreinte (la dernière étude de ce travail de thèse portant sur une population d'individus présentant un trouble de personnalité borderline).

En préambule à cette deuxième étude, nous avons traduit en français l'échelle *Five-Factor Borderline Inventory-Short Form* (FFBI-SF, Nasello et al., 2021b) afin de mesurer les traits de personnalité borderline. La traduction présente de bonnes propriétés psychométriques en termes de validité, de fidélité et de structure factorielle.

Toutefois, cette échelle n'est pas un outil diagnostique, mais elle mesure le degré de présence des caractéristiques propres au trouble borderline de manière dimensionnelle (voir Annexes C & D pour davantage des informations détaillées).

Récapitulatif de l'étude 2 :

Une fois cette adaptation validée et publiée, nous l'avons utilisée comme matériel pour notre deuxième étude. Cette seconde étude cherchait : (1) à tester la validité du DPM en étudiant les traits de personnalité borderline et l'empathie pour en déterminer leur pouvoir de prédiction sur les choix utilitaires au sein de la population générale, (2) à comparer les choix utilitaires via deux matériels de type trolley (un scénario *causing harm* : le dilemme du chirurgien ; et un scénario *causing inconvenience* : le dilemme de l'entreprise ; voir Annexe E) et (3) à implémenter une nouvelle perspective dans les scénarios (i.e., celui parmi les cinq) pour déterminer si cela engendrait des changements significatifs au niveau des choix orientés vers le groupe (i.e., choix utilitaires).

Article: Moral decision-making in trolley problems and variants: how do participants' perspective, borderline personality traits, and empathy predict choices?

THE JOURNAL OF PSYCHOLOGY
https://doi.org/10.1080/00223980.2023.2206604

Moral Decision-Making in Trolley Problems and Variants: How Do Participants' Perspectives, Borderline Personality Traits, and Empathy Predict Choices?

Julian A. Nasello[a,b], Benoit Dardenne[b], Michel Hansenne[b], Adélaïde Blavier[b] and Jean-Marc Triffaux[a,b]

[a]Psychiatric Day Hospital "La Clé"; [b]University of Liège

ABSTRACT

The aim of the present study was to demonstrate and extend the causal effect of participants' perspectives on moral decision-making using trolley problems and variants. Additionally, we investigated whether empathy and borderline (BDL) personality traits predicted participants' choices in these scenarios. We used both a classical trolley problem (a causing harm scenario) and an everyday trolley-like problem (a causing inconvenience scenario). Participants ($N = 427$, women: 54%) completed BDL traits and empathy questionnaires and, randomly, the two types of trolley problems, presenting both three different perspectives. Our study provided strong evidence that the perspective from which participants were enrolled in the trolley problem caused significant changes in their moral decision-making. Furthermore, we found that affective empathy and BDL traits significantly predicted participants' decisions in the causing inconvenience scenario, while only BDL traits predicted choices in the causing harm scenario. This study was original in providing new experimental materials, causal results, and highlighting the significant influence of BDL traits and affective empathy on moral decision-making. These findings raised fundamental questions, which are further developed in the discussion section.

ARTICLE HISTORY
Received 1 April 2022
Accepted 18 April 2023

KEYWORDS
Moral judgment; moral development; perspective-taking; psychopathology; personality

Introduction

Moral dilemmas have been studied by social psychology, philosophy, and economics for decades, notably because they have several societal implications (e.g., in public health or politics). Recently, they were at the heart of preoccupations during the COVID-19 pandemic, given that many health centers were overwhelmed and had to make unusual ethical choices (e.g., choosing who would benefit from a respirator).

One of the most famous moral dilemmas is the trolley problem. Initially, Foot (1967) and, a few years later, Thomson (1976) developed a problem presented as follows: imagine that you are a driver of a runaway tram that can only steer from one track to another. Five workers are standing on one track, and one workman on the other. The only option is to divert a lever to switch the tramway from one track

CONTACT Julian A. Nasello julian.nasello@gmail.com Psychiatric Day Hospital "La Clé", Boulevard de la Constitution 153, Liège, Belgium.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00223980.2023.2206604.

to another, which leads to choosing between killing five people and saving one; or killing one to save the five (Foot, 1967). On the other hand, Thomson's variant (1976) places you on a bridge where an obese workman stands next to you. On this bridge, you see that a runaway tramway will hit and kill the five workers on the track if nothing is done. Hence, you have to choose between pushing to death the obese workman on the track to stop the tramway and save the five; or doing nothing and letting the five get killed.

Greene and collaborators popularized these problems in 2001 with their publication in *Science*. Several studies demonstrated that participants' choices were mainly group-oriented decisions (GOD, also called *utilitarian* choices: choosing to save the five). Indeed, around 90% of people decided to kill the one to save the five in Foot's version. However, this tendency decreased drastically in Thomson's variant (roughly 10% decided to push the obese worker: Hauser et al., 2007).

Greene et al. (2001, 2004) and Greene (2007) classified Foot and Thomson's versions and developed the Dual-Process Model to understand these drastic decision-making changes. Indeed, Foot's version has been classified as an *impersonal* scenario, and Thomson's variant as a *personal* scenario. They have argued that impersonal moral scenarios involve more cognitive processes and fewer social-emotional responses, leading to higher GOD (i.e., *utilitarian* choices). However, personal scenarios involve the opposite processes (i.e., it elicits higher social-emotional responses and lower cognitive processes), leading to fewer GOD. The authors named this model the *Dual-Process Model* (DPM, Greene et al., 2001, 2004, 2008).

Critics of Classical Trolley Problems and Solutions

Bauman et al. (2014) raised several limits on trolley problems. They reported that participants tend to be amused rather than concerned and described trolley problems as neither immersive nor realistic, too far from daily-life activities. Indeed, killing someone is fortunately far from daily-life activities (Nasello et al., 2021), and several authors have attempted to overcome these trolley problems' limits. For instance, Navarrete et al. (2012) proposed a virtual reality study of the trolley problem, immersing participants in a virtual environment. More recently, several authors proposed to overcome Bauman et al.'s limits (2014) by using some everyday trolley-like problems that are more realistic and plausible (Lotto et al., 2014; Nasello et al., 2021; Pletti et al., 2017; Takamatsu, 2019; Vyas et al., 2017). These ecologically valid variants showed more generalizable findings, which is of substantial interest when studying psychopathological populations. Therefore, to display choice differences in both scenario types, this study used an everyday trolley-like problem and a classical variant based on Thomson's scenario (i.e., the transplant dilemma). Bruers and Braeckman (2014) considered Thomson's scenario and the transplant dilemma similar.

Participant's Perspective on Trolley Problems and Variants: A Broader Understanding?

An important element in trolley problems and variants has been depicted in a recent study. This study demonstrated a causal effect of participants' perspectives on everyday

trolley-like problems (Nasello et al., 2021). The authors created two versions of a similar trolley-like problem and invited participants to be alternatively two different protagonists of these scenarios. In one perspective, the participant was asked to be a driver of a university tramway who had to decide whether to stop or not to board one tardy student. This scenario is comparable to impersonal scenarios from classic trolley problems (they called it "*a no-direct impact scenario*" because the participant's decision will not impact him/her directly). In the other one, the participant was a tardy student (they called it "*a direct impact scenario*" because the participant's decision will impact him/her directly: attending or not the lecture). As expected, the authors showed a higher ratio of GOD in the driver's perspective (75-25), and this proportion decreased when the participant was the tardy student (50-50).

In the present study, we targeted a missing element by inviting the participant to enroll a protagonist in a third perspective: the one among the five. Hence, we aimed to replicate Nasello et al.'s experimental design (2021) and test whether an additional perspective (i.e., Perspective C, see Figure 1) would impact moral decision-making to deepen the results previously obtained.

The Role of Emotional Processes and Empathy in Trolley Problems and Variants

Based on the DPM (Greene et al., 2001), several authors focused on the role of emotional processes. For instance, Choe and Min (2011) showed that anger, disgust, and empathy significantly predict utilitarian judgments. Anger increased GOD levels, while empathy and disgust decreased these levels. Empathy is "an innate ability to perceive and be sensitive to the emotional states of others, coupled with a motivation to care for their well-being" (Smith et al., 2017, p. 2) and is usually composed of two domains: affective and cognitive empathy. The former describes the ability to feel what others feel, and the latter refers to the capacity to comprehend others' perspectives and put oneself in someone's shoes. Several authors found that (1) a lack of affective empathy characterized psychopathic traits (Burghart & Mier, 2022) and (2) that this affective empathy depletion predicted GOD significantly in trolley problems (Decety & Cowell, 2014; Gleichgerrcht & Young, 2013; Greene et al., 2004; Patil & Silani, 2014; Takamatsu, 2018). While these findings are important, we have no experimental results on contrasted psychopathology (i.e., clinical populations presenting higher affective empathy scores).

Empathy is "*an innate ability to perceive and be sensitive to the emotional states of others, coupled with a motivation to care for their well-being*" (Smith et al., 2017, p.2) and is usually composed of two domains: affective and cognitive empathy. The former describes the ability to feel what others feel, and the latter refers to the capacity to comprehend others' perspectives and put oneself in someone's shoes.

As mentioned, several authors found that (1) a lack of affective empathy characterized psychopathic traits (Burghart & Mier, 2022) and (2) that this affective empathy depletion predicted GOD significantly in trolley problems (Decety & Cowell, 2014; Gleichgerrcht & Young, 2013; Greene et al., 2004; Patil & Silani, 2014; Takamatsu, 2018). While these findings are important, we have no experimental results on contrasted psychopathology.

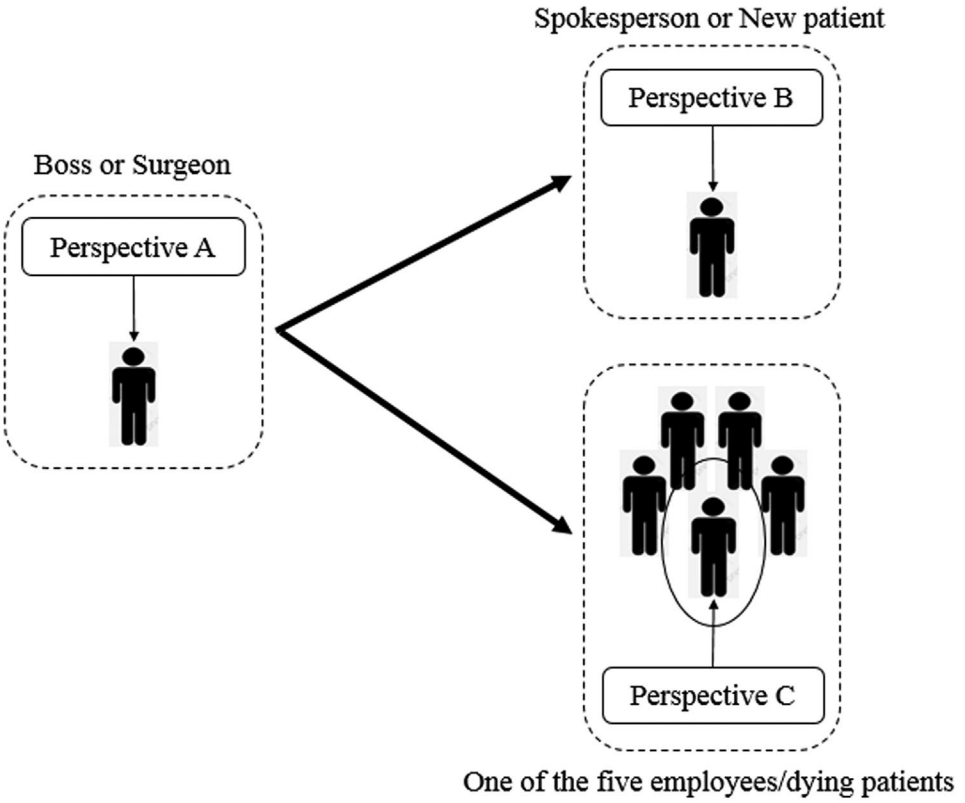


Figure 1. Decision-maker's perspectives on trolley problems and variants.
Note. This figure depicts the three perspectives enrolled by the participant (i.e., the decision-maker) for the causing harm and inconvenience scenarios.

The Opposite Side of the Affective Empathy Continuum

Compared to psychopathic traits, borderline personality traits are characterized by salient divergences in affective empathy and emotional functioning. Furthermore, a pervasive pattern of instability of affects, interpersonal relationships, self-image, and marked impulsivity characterizes borderline (BDL) personality disorders. This personality disorder is a severe and persistent psychiatric disorder with a lifetime prevalence estimated to be between 1 and 3% in the general population (Leichsenring et al., 2011).

Therefore, despite some shared symptoms (e.g., marked impulsivity), psychopathy and BDL personality disorders are fundamentally different in their emotional semiology: while psychopathy is characterized by emotional callousness (Hare, 1996; Hare & Neumann, 2009) and emotional hyporesponsiveness compared to BDL (Herpertz et al., 2001), BDL personality disorders are instead marked by heightened emotional reactivity (Crowell et al., 2009). Several authors considered that affective and emotional intensity and lability are central symptoms of BDL (Richetin et al., 2017; Southward & Cheavens, 2018), regardless of age (Peckham et al., 2020). Furthermore, in comparison with healthy control populations, a recent systematic review showed that BDL

presents higher affective empathy levels and a deficiency in perspective-taking (Salgado et al., 2020), while another review (Burghart & Mier, 2022) showed that psychopaths display a lack of emotional/affective empathy (as sustained by several findings: Blair, 2005; Hare, 1993; Robinson & Rogers, 2015; Sandoval et al., 2000; Van Dongen, 2020) and a deficiency in perspective-taking as well.

Therefore, based on the differences between psychopathic and borderline personality traits and their impact on emotional functioning in moral decision-making, our study aims to investigate whether subclinical borderline personality traits predict a reversed pattern of moral decision-making in trolley problems and variants, as opposed to the effect of psychopathic traits. The underlying concept is that emotional processes significantly influence moral judgments, as stated by Greene et al. (2001). Hence, affective empathy—an emotional process that amplifies or attenuates the emotional experience of the protagonists in moral dilemmas—or psychopathological personality traits characterized by higher or lower affective empathy levels should influence moral decision-making.

Objectives and Hypotheses

This study had two principal purposes: (1) demonstrating and extending the causal effect of participants' perspectives on moral decision-making and (2) determining whether higher affective empathy and borderline traits significantly predict lower GOD in trolley problems and variants. All hypotheses have been depicted in Figure 2.

Like Nasello et al. (2021), we aimed to demonstrate that participants' perspectives in trolley scenarios cause significant changes in their moral decision-making. In the present study, we introduce a third perspective (i.e., Perspective C) to complement perspectives A and B (see Figure 1) already used by these authors. To generalize this

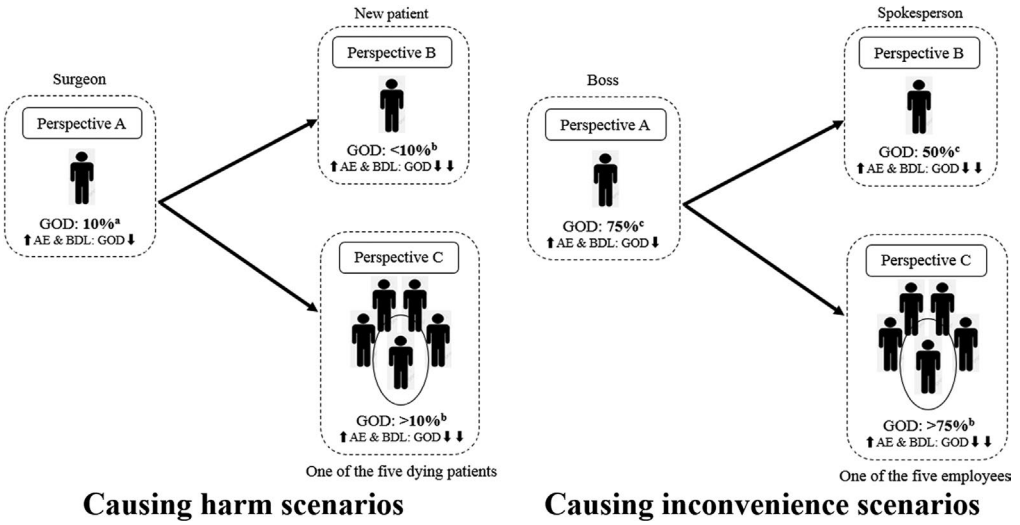


Figure 2. Hypotheses.
Abbreviations. GOD = Group-Oriented Decision; AE = Affective Empathy; BDL = Borderline traits. [a]Hauser et al. (2007). [b]Greene et al. (2001). [c]Nasello et al. (2021).

causal impact, we used two types of scenarios: a *causing harm* scenario (i.e., one or five protagonist(s) will die according to the participant's decision) and a *causing inconvenience* scenario (i.e., one or five protagonist(s) will get fired according to the participant's decision). Following the predictions made from the DPM (Greene et al., 2001), the *causing harm* scenario should display lower GOD levels than the *causing inconvenience* scenario. Note that we chose GOD as a reference because it was the common option in the three perspectives of both scenarios. Therefore, lower GOD involved more individual-oriented or self-oriented decisions and vice versa.

In the company's dilemma (i.e., the causing inconvenience scenario), we hypothesized that Perspective A would present higher GOD levels (around 75%, like Nasello et al.'s findings, 2021) and Perspective C even higher GOD levels. However, Perspective B would present lower GOD levels (around 50%, like Nasello et al.'s findings, 2021). In the transplant dilemma, we expected around 10% of group-oriented decisions (GOD) in Perspective A, as found by Hauser et al. (2007) and as predicted by the DPM (Greene et al., 2001). Note that Perspective A is the only perspective tested in trolley dilemmas' scientific literature. For other perspectives, we hypothesized that Perspective B would present lower GOD and Perspective C the highest GOD levels: depending on which perspective the participant is, s/he will tend to favor his/her position if the consequences of the decision will directly impact him/her (see Figure 2 for an illustration).

Second, based on the predictions made from the DPM (i.e., higher social-emotional responses elicited by dilemmas lead to lower GOD levels), we hypothesize that higher affective empathy and higher BDL traits will predict lower GOD levels in trolley problems and variants. The prediction made for BDL traits lies on two main assumptions: (1) BDL people are characterized by higher affective empathy levels, according to a recent systematic review (Salgado et al., 2020), and higher levels of affective empathy are expected to decrease GOD levels on decision-making. (2) BDL people present the opposite emotional semiology compared to psychopaths (a population displaying higher GOD levels). Therefore, because BDL people are more prone to experience intense and unstable emotions, we sought to determine whether higher BDL traits (and affective empathy) would significantly predict less GOD in trolley problems and variants. Based on the DPM, we can speculate that higher affective empathy and BDL traits would significantly predict lower GOD levels in Perspectives B and C than in Perspective A (see Figure 2).

Method

Participants & Procedure

Four hundred and twenty-seven participants voluntarily took part in our study ($N = 427$; $n_{women} = 231$; $M_{age} = 25.3$; $SD_{age} = 5.10$, see Table 1). The majority of participants were Caucasians (88%), unmarried (77%), and had completed at least a Bachelor's degree (60%). Participants were required to have no history of psychiatric or neurological disorders to be eligible to participate. They were recruited online through social media advertisements and were not offered any financial compensation for their participation. Participants completed three questionnaires (i.e., a demographic questionnaire, the

Table 1. Descriptive Statistics.

Variables	*N*	*Mean*	*SD*	*Min*	*Max*	*ω*
Age	427	25.3	5.10	18	38	
Women	231	25.4	4.84	18	38	
Men	196	25.3	5.40	18	38	
Empathy						.85
Affective empathy	427	41.3	7.27	14	55	.81
Women	231	44.4	6.13	17	55	
Men	196	37.6	6.75	14	55	
Cognitive empathy	427	36.8	4.51	20	45	.81
Women	231	37.9	4.05	22	45	
Men	196	35.5	4.67	20	45	
Borderline traits						
Total	427	110	27.7	54	188	.94
Women	231	112	27.7	54	188	
Men	196	109	27.8	54	186	

Basic Empathy Scale, and the Five-Factor Borderline Inventory, Short Form) before being presented with two types of trolley problems, each containing three perspectives (i.e., A, B, and C, as shown in Figure 1) in random order. Sample size was determined using G*Power 3.1.9.7 (Faul et al., 2007), with standard criteria set at $\alpha = 0.05$ and power = 0.80, requiring a minimum of 372 participants.

The Research Ethics Committee of the Psychology Department at the University Liège (Belgium) approved this research, Ref.: 1920-92

Materials

Empathy

The Basic Empathy Scale (Jolliffe & Farrington, 2006; French version: D'Ambrosio et al., 2009) is a 20-item scale assessing two empathy components (i.e., affective and cognitive empathy). Affective empathy refers to *the capacity to experience the emotions of another* (Bryant, 1982) and is composed of 11 items (e.g., *I get caught up in other people's feelings easily*). On the other hand, cognitive empathy is defined as *the capacity to comprehend the emotions of another* (Hogan, 1969) and is composed of 9 items (e.g., *I can often understand how people are feeling even before they tell me*). Items are rated on a 5-point Likert scale from 1 (*It does not describe me very well*) to 5 (*It describes me very well*). The scale displays good internal reliability (present study: affective empathy: ω = .81; cognitive empathy: ω = .81; total empathy: ω = .85).

BDL Traits

We used the 48-item self-rated Five-Factor Borderline Inventory, Short Form (FFBI-SF, DeShong et al., 2016; French version: Nasello et al., 2023) to measure subclinical borderline traits. Each item is rated on a 5-point Likert scale, from 1 (*Strongly disagree*) to 5 (*Strongly agree*) (e.g., "*I often conflict with people that are close to me*" or "*My emotions can get out of control*"). In this study, we based our analyses on the total BDL score. Nasello et al. (2023) reported that people scoring ≥162 could be considered as presenting salient BDL traits. The FFBI-SF displays good internal reliability (present study: ω = .94).

Trolley Problems and Variants

Two types of trolley problems, each with three perspectives, were used to assess moral decision-making (i.e., Perspectives A and C: GOD versus Individual-Oriented Decisions, Perspective B: GOD versus Self-Oriented Decisions). Note that Perspective C involves a mixed decision between GOD and Self-Oriented Decisions. As a classical trolley problem, the transplant dilemma is a *causing harm* scenario. In this scenario, a brilliant surgeon has five patients suffering from a specific organ failure, and they will die without a transplant. Another patient enters the hospital because s/he needs urgent but benign surgery. This surgeon notices that this new patient's organs are all compatible with the five dying patients. The participant is presented with two options: (1) perform a lethal maneuver on the new patient during their surgery to save the five dying patients (the new patient will die, but the other five will live); (2) perform the new patient's surgery according to best practices to save his/her life (the new patient will live, but the other five will die). The participant is randomly assigned to be the surgeon, the new patient, or one of the five dying patients.

The second dilemma was a *causing inconvenience* scenario, named the company's dilemma. The general parameters of this everyday trolley-like problem are that the participant is a worker of a small company (i.e., the boss, the spokesperson, or one of the five employees) and, because of the current pandemic situation, this company has to downsize its workforce. Panicked, the six employees have decided to go on strike to claim employment security and termination indemnity in case of layoff. To calm down the situation, the boss has called for a meeting with the spokesperson of the employees. In this meeting, the boss tries to understand the spokesperson's situation and explains that s/he does not have several options. In the three perspectives, the boss proposes a two-option arrangement behind closed doors (with immediate effects and no consultation meeting with the other employees). The first option is to contractually protect the spokesperson's job but lay off the other five employees without indemnity termination. The other option is to lay off the spokesperson without indemnity termination but to contractually protect the jobs of the other five employees[1].

Note that in all scenarios, the decision-maker makes the decision alone; there is no joint decision between the participant and the other protagonists. Both scenarios were primed using the COVID-19 pandemic to enhance their immersive quality (see Supplementary Materials for a complete description). In Perspectives A and C, participants were presented with same-gender scenarios, with the target being a man when the participant was a man and a woman when the participant was a woman.

Data Analyses and Experimental Design

The principal analyses were performed using a Generalized Mixed Model (GMM) to determine whether perspectives, empathy, and BDL traits significantly predict moral decision-making in trolley problems and variants. For this purpose, we run two GMMs separately: one for *causing harm* scenarios and one for *causing inconvenience* scenarios. There were three continuous independent variables (i.e., BDL traits, affective and cognitive empathy) and *Perspective* as predictors of the binary choices (GOD were coded "1" and other choices "0", i.e., individual-oriented decisions or self-oriented decisions). Intercept and ID were entered as random effects, and all continuous

predictors were centered. See Gallucci (2019) for a complete description of the GMM procedure. Because of significant differences between men and women, gender was introduced in the GMM for *causing harm* scenarios but not for *causing inconvenience* scenarios. As descriptive statistics, chi-square tests highlighted choice differences in the two types of scenarios (x 3 perspectives).

The analyses were conducted using the JAMOVI computer software, version 1.6.23 (The Jamovi Project, 2019). The raw data file can be found at the following DOI link: https://osf.io/t4h7f/.

Results

Descriptive Statistics

Causing Inconvenience Scenarios

As expected, participants made higher GOD levels (i.e., laying off the spokesperson) in Perspectives A and C ($\mathbf{P}_A$: 92% of GOD; $\mathbf{P}_C$: 94% of GOD; see Figure 3a). There was no significant choice difference between these two perspectives ($\chi^2_{(1)} = 0.081$; $p = .776$). Also in line with our expectation, the GOD frequency significantly dropped down in Perspective B (54% of GOD) compared to Perspectives A and C ($\mathbf{P}_{Avs.B}$: $\chi^2_{(1)} = 41.03$; $p < .001$; $\mathbf{P}_{Cvs.B}$: $\chi^2_{(1)} = 44.7$; $p < .001$). There was no difference between men and women in the three perspectives ($\mathbf{P}_A$: $\chi^2_{(1)} = 1.08$; $p = .299$; $\mathbf{P}_B$: $\chi^2_{(1)} = 0.087$; $p = .769$; $\mathbf{P}_C$: $\chi^2_{(1)} = 3.38$; $p = .066$).

Causing Harm Scenarios

As expected, the GOD levels (i.e., killing the new patient) were low in the three perspectives ($\mathbf{P}_A$: 9% of GOD; $\mathbf{P}_B$: 13% of GOD; $\mathbf{P}_C$: 18% of GOD; see Figure 3b). Furthermore, there were significant differences between Perspectives B and C ($\mathbf{P}_{Cvs.B}$: $\chi^2_{(1)} = 4.36$; $p = .037$) and Perspectives A and C ($\mathbf{P}_{Avs.C}$: $\chi^2_{(1)} = 13$; $p < .001$), but we did not found any significant difference between Perspectives A and B ($\mathbf{P}_{Avs.B}$: $\chi^2_{(1)} = 2.42$; $p = .120$). The GOD level of Perspective B was unexpected. Finally, we found significant gender differences in Perspectives A and C ($\mathbf{P}_A$: $\chi^2_{(1)} = 7.45$; $p = .006$; $\phi = .132$; $\mathbf{P}_C$: $\chi^2_{(1)} = 7.92$; $p = .005$; $\phi = .136$). In Perspective A, men made 13% of GOD and women 6%, and men made 24% of GOD and women 13% in Perspective C. In both perspectives, men made more than twice GOD as women.

Generalized Mixed Model Analyses

Causing Inconvenience Scenarios (See Supplementary Materials for the Correlation Matrix)

The GMM ($\chi^2/df = 0.836$; $R^2_{Marg.} = .308$; $R^2_{Cond.} = 0.373$) showed that "Perspective" was a strong significant predictor of decision-making ($\mathbf{P}_{C\text{-}B}$: $B = 2.709$; $exp(B) = 15.02$; $p < .001$; $\mathbf{P}_{A\text{-}B}$: $B = 2.42$; $exp(B) = 11.25$; $p < .001$; see Table 2). When moving from Perspective B to C, the odds of making a GOD were multiplied by 15, whereas when moving from Perspective B to A, the odds were multiplied by 11.3. Furthermore, BDL traits significantly and negatively predicted GOD ($B = -0.01$; $exp(B) = 0.992$; $p = .009$; see Table 2): each additional point of BDL traits decreased by 1% the odd to make a

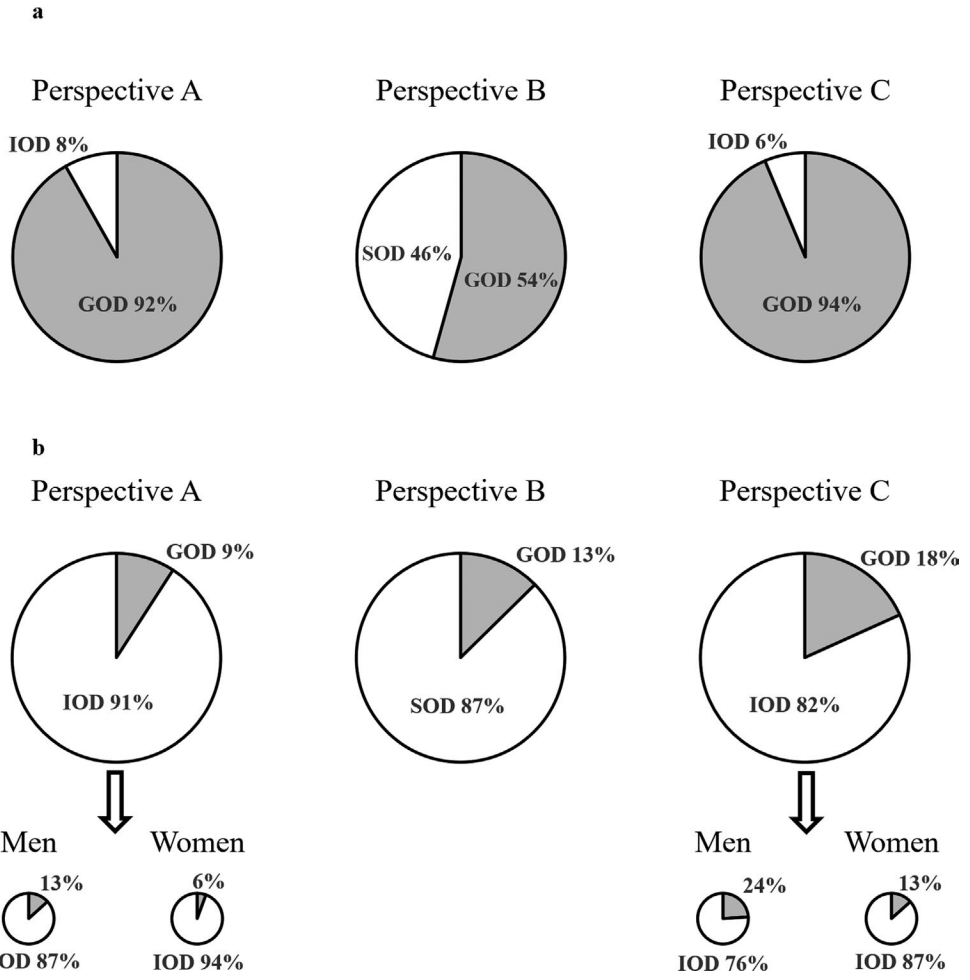


Figure 3. (a) Causing inconvenience scenarios, (b) Causing harm scenarios.
Note. Diagrams displaying men and women's choices were represented when a significant gender difference was found.
Abbreviations: GOD = Group-Oriented Decision; SOD = Self-Oriented Decision; IOD = Individual-Oriented Decision.

Table 2. Coefficients of the General Mixed Model Analysis.

					95% Exp(B) CI			
	Names	E	SE	Exp(B)	Lower	Upper	z	p
CI	Intercept	1.902	0.130	6.70	5.20	8.64	14.7	< .001
	$Perspective_{C-B}$	2.71	0.251	15.02	9.19	24.6	10.8	< .001
	$Perspective_{A-B}$	2.42	0.231	11.3	7.15	17.7	10.5	< .001
	Affective empathy	0.037	0.014	1.038	1.01	1.07	2.74	0.006
	Cognitive empathy	0.013	0.022	1.013	0.970	1.06	0.573	0.567
	BDL Traits	−0.008	0.003	0.992	0.985	0.998	−2.61	0.009
CH	Intercept	−3.11	0.298	0.045	0.001	0.071	−10.5	< .001
	Gender	−0.835	0.337	0.434	0.165	0.961	−2.48	0 .013
	$Perspective_{C-B}$	1.15	0.261	3.151	1.83	10.2	4.40	< .001
	$Perspective_{A-B}$	0.514	0.266	1.673	1.053	5.35	1.94	0.053
	Affective empathy	−0.007	0.025	0.993	0.924	1.06	−0.284	0.776
	Cognitive empathy	−0.062	0.038	0.940	0.872	1.03	−1.63	0.102
	BDL Traits	0.013	0.006	1.013	1.001	1.02	2.29	0.022

Abbreviations: CI = Causing Inconvenience; CH = Causing Harm; BDL Traits = Borderline Traits.

GOD. Finally, affective empathy was a positive and significant predictor of GOD ($B = 0.037$; $exp(B) = 1.038$; $p = .006$; see Table 2): each additional affective empathy point increase of 4% the odd to make a GOD.

Graphical plots showed significant predictive effects of affective empathy and BDL traits in Perspective B (see Figure 4A and B). However, cognitive empathy did not significantly predict GOD ($B = 0.013$; $p = .567$; see Table 2).

Causing Harm Scenarios

The GMM ($\chi^2/df = 0.381$; $R^2_{Marg.} = .0896$; $R^2_{Cond.} = 0.583$) displayed a significant predictive effect of "Perspective" on GOD ($\mathbf{P}_{C\text{-}B}$: $B = 1.15$; $exp(B) = 3.15$; $p = < .001$; see Table 2): changing from Perspective B to C multiplied by 3.2 the probability to make a GOD (i.e., killing the new patient to save the five). BDL traits also significantly predicted (in a positive way) GOD ($B = 0.013$; $exp(B) = 1.013$; $p = .022$; see Table 2),

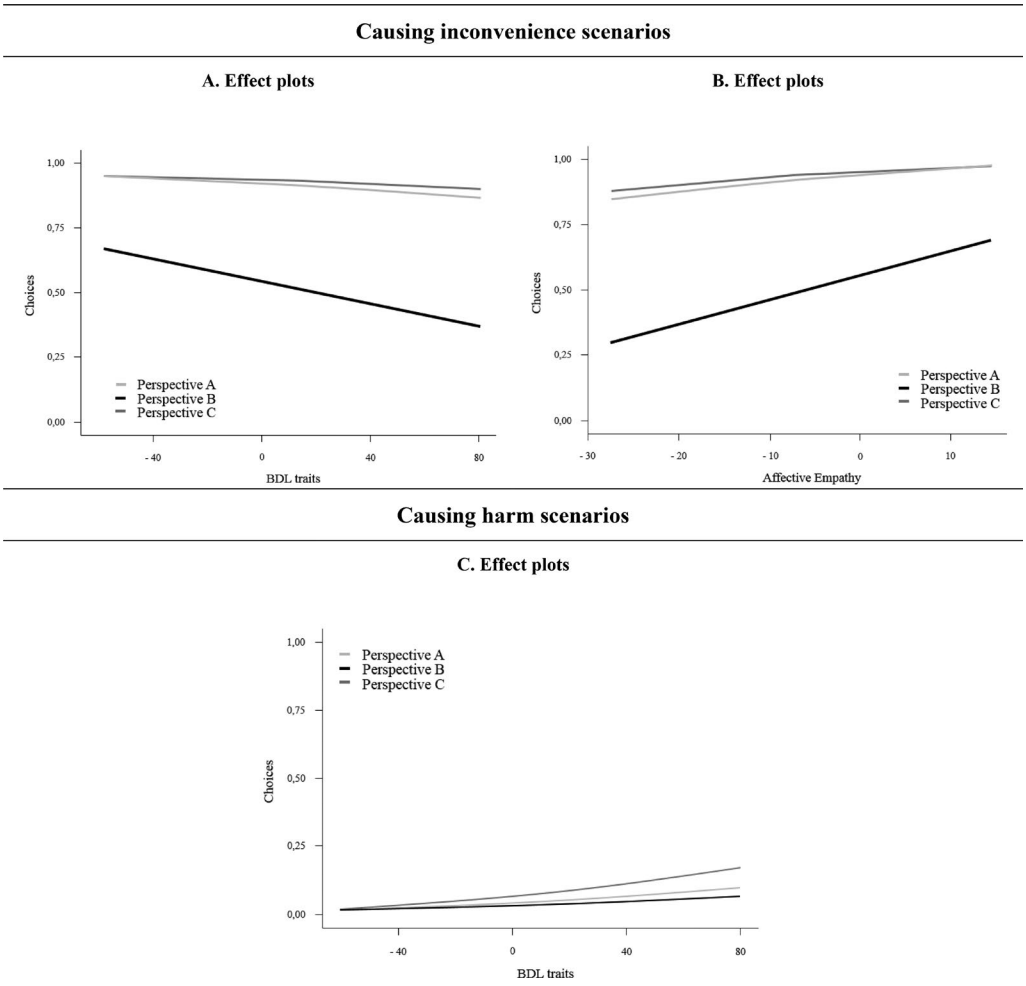


Figure 4. GMM effect plots for causing inconvenience and causing harm scenarios.
Note. The X-axis displays BDL traits (A & C) and affective empathy (B), while the Y-axis depicts participants' choices (the closer to 1, the more GOD). Each separate line represents the three perspectives.

meaning that each additional point on the FFBI-SF increased by 1% the odd to make a GOD. This effect was especially true for Perspective C (i.e., when the participant is one of the five dying patients; see Figure 4C). Gender significantly predicted GOD ($B = -0.835$; $exp(B) = 0.434$; $p = .013$; see Table 2). Surprisingly, neither affective empathy ($B = -0.007$; $p = .776$) nor cognitive empathy ($B = -0.062$; $p = .102$; see Table 2) showed a significant predictive effect on GOD.

Discussion

The Role of the Decision Maker's Perspective on Trolley Problems and Variants

We have replicated and extended Nasello et al. (2021) findings, demonstrating that the participant's perspective has a causal effect on moral decision-making in causing harm and inconvenience scenarios.

In causing inconvenience scenarios, we replicated Nasello et al. (2021) findings for Perspective B, but we found higher levels of GOD for Perspective A (92% vs. 75%). This discrepancy might be explained by introducing a notion of probabilities to get the lecture for a certain amount of passengers in Nasello et al. (2021) findings. The notion of uncertainty affecting a few protagonists seems to lower the propensity to make some GOD. For instance, Kahneman (2013) and Kahneman and Tversky (1982) showed that uncertain events influence decision outcomes. However, our GOD level for Perspective A in the causing inconvenience scenario is very close to previous findings using Foot's trolley problem (e.g., Navarrete et al. (2012) reported 90.5% of GOD), suggesting that moral intensity is similar in these two kinds of dilemmas. Despite the major differences in the consequences involved in these moral dilemmas (i.e., someone who dies vs. someone who loses his/her job), the moral agent (i.e., the person who makes the moral decision; Jones, 1991) does not intentionally perform an action that violates moral norms, unlike in Thomson's variant (see below). Jones (1991) stressed that *intention* has a central role in moral decision-making and relied on Fishbein and Ajzen's theory (1975), stating that intention is the best predictor of a person's behavior. Looking at the perspectives in detail, we found that Perspectives A and C presented the same GOD levels, suggesting that the common good interest (the aggregate welfare) is maximized in causing inconvenience conditions as long as people do not have to choose between others and themselves. In that specific case (Perspective B), people were more balanced, and those who made self-oriented choices probably placed higher values on themselves (Jones, 1991) by consolidating the post-decision option using rational elements (e.g., family duties; see Svenson, 1992, 1996).

The situation was utterly different in the causing harm scenarios. The overall GOD levels drastically decreased compared to the causing inconvenience scenarios, aligning our predictions. As mentioned earlier, the moral violation is extremely high in the transplant dilemma because participants must voluntarily perform a lethal maneuver to save the five, demonstrating that intentions played a crucial role (Jones, 1991). Hunt and Vitell (1986) have also perfectly demonstrated that people perceive the factual reality of one ethical problem, assess the inherent righteousness of a decision, and the amount of good or bad embodied in the consequences of that decision. All of these stages, and their combination, play an essential role in moral decision-making.

In this scenario, we replicated the GOD rate (around 10%) obtained by Hauser et al. (2007) in Perspective A, and we found the highest GOD levels in Perspective C, as expected (see Figure 2). However, contrary to our expectations, we did not find the lowest GOD rate in Perspective B compared to Perspectives A and C. The result is surprising because, compared to other perspectives, a substantial number of people would sacrifice themselves to save the five dying patients. To explain this unexpected result, it is possible that some individuals may have considered this situation more like assenting the doctor to commit assisted suicide than murder. This element may have slightly softened the perceived seriousness of the situation, explaining why we obtained a higher percentage than expected (13% instead of <10%). When analyzing the participants who sacrificed themselves in perspective B of the transplant dilemma, we discovered that most were young Caucasian bachelor students (Age: $Mdn = 22$; $SD = 5{,}22$). The reason why this small segment of the population is more inclined to make this type of decision remains to be determined. For instance, an unmeasured dispositional trait (e.g., extreme altruism or social desirability propensity) might have played a significant role in this perspective[2].

Finally, we found some gender differences in causing harm scenarios: men were twice as likely as women to make GOD in Perspectives A and C. Gender differences have already been displayed in the scientific literature in causing harm scenarios (Fumagalli et al., 2010a, 2010b; Friesdorf et al., 2015; Armstrong et al., 2019), but they are nuanced when participants embody the new patient's perspective (i.e., Perspective B; in this perspective, men and women showed similar GOD levels). As Nasello et al. (2021) mentioned, these differences are presumably due to stronger affective responses to harming someone for women and more cognitive evaluations of outcomes for men.

Overall, we replicated and extended previous findings in this study, demonstrating a causal effect of the decision maker's perspective on trolley problems and variants. Our results highlight the theoretical significance of considering the variable "perspective" in the field of moral decision-making. Specifically, being enrolled in the perspectives of other protagonists forced individuals to balance their personal interests with those of others in a particular situation or scenario, leading to significant changes in moral decision-making.

BDL Traits as a Predictor of GOD Levels

We assumed that higher levels of affective empathy and BDL traits would predict less GOD (see Figure 2), in accordance with the predictions made from the DPM (Greene et al., 2001). The present study's findings partly confirmed this assumption. Our findings showed that affective empathy and BDL traits significantly predict GOD under specific conditions but not under all perspectives, as we expected.

For causing inconvenience scenarios, affective empathy and BDL traits were significant GOD predictors in Perspective B. However, contrary to our expectations, they did not similarly predict GOD. When the participant was the spokesman, the more affective empathy, the more GOD (this result does not align with the DPM prediction); inversely, the more BDL traits, the less GOD (this result aligns with the DPM prediction). On the other hand, only BDL traits significantly predicted GOD in Perspective C of causing harm

scenarios. Participants with higher BDL traits were more likely to make GOD when they were among the dying patients (this result does not align with the DPM prediction).

These findings suggest subtle egocentric choices for individuals with higher BDL traits. We argue that this tendency is subtle because (1) participants' decisions were made in contexts where other protagonists could not determine who had made the decision (see Supplementary Materials for a complete description); and (2) because, as previously mentioned, Perspective C of causing harm scenarios is a mixed decision between GOD and self-oriented decisions.

Our results displayed that GOD increased as a function of BDL traits when the participant was the spokesperson, while higher BDL traits increased the likelihood of making a GOD when s/he was one of the five dying patients. These results: (1) support Saunders et al.'s findings (2015), which demonstrated that people suffering from BDL present selectively impaired altruism, and (2) suggest that BDL individuals may be characterized by a higher propensity to make egocentric choices in specific moral situations.

Furthermore, this subtle propensity for egocentric choices may affect their interpersonal difficulties and unstable relationships. To explain this tendency, individuals with higher BDL traits may be more sensitive to moral intensity (a concept developed by Jones, 1991) or moral violation, leading them to make different moral decisions. Importantly, it should not be assumed that favoring the group (GOD) is the 'right' decision. From perspective B, we can argue that a GOD is an altruistic decision because participants favor the group instead of themselves. However, a GOD is not necessarily the right decision or the decision that should be favored. Like in classical trolley problems, the same underlying question is present in this perspective: why should the jobs of five coworkers be considered more valuable than one's own, and vice versa?

Is Empathy a Relevant Predictor of GOD Levels?

Interestingly, affective empathy had the opposite effect of BDL traits in predicting moral decision-making in Perspective B of scenarios involving inconvenience. This finding was unexpected, given the predictions made from the DPM (Greene et al., 2001). The DPM sustains that emotional processes play a significant role in moral decision-making (particularly in personal or direct-impact scenarios such as those in Perspectives B and C), where individuals are expected to make fewer GOD decisions (and this propensity was expected to be even higher for people presenting higher levels of affective empathy). However, our study found that affective empathy linearly predicted more GOD decisions in Perspective B involving inconvenience scenarios (see Figure 4B). This result was more consistent with the Empathy-Altruism Hypothesis proposed by Batson (1987, 2011). The Empathy-Altruism Hypothesis assumes that empathic concern (an affective empathy domain) produces altruistic motivation, and this hypothesis was supported by a 35-year review of empirical research (Batson et al., 2015). The authors demonstrated that empathy leads to more sensitive and constant help toward others (Batson et al., 2015), and this appears to be especially true when people are directly involved in causing inconvenience situations (like in Perspective B). In parallel to Batson's claims, Decety and Cowell (2015) showed that affective empathy (empathic concern) has evolved to favor ingroup members and can bias decision-making by valuing one individual over a group.

However, our results showed only a small overall predictive effect of affective empathy, similar to BDL traits, which suggests that empathy is not the most relevant predictor of moral judgments and decision-making (Decety, 2021; Decety & Cowell, 2015). These findings raise two fundamental questions: (1) to what extent do emotional processes affect moral decision-making as proposed by the DPM? (2) why did cognitive processes such as cognitive empathy not significantly predict moral decision-making in our regression models?

Conclusion

As indicated by our results, while the DPM (Greene et al., 2001) provides a good explanation for certain outcomes, it appears to (1) overestimate the role of emotional processes, (2) generate predictions that do not entirely align with empirical findings (as also shown by Smillie et al., 2021), and consequently (3) offer a poor explanation of moral decision-making.

Given the limited influence of affective empathy on decision-making in this study and prior research (e.g., Nasello et al., 2021), we agree with Horne and Powell's argument (2016) that the presumed strong connection between emotion and moral judgment or decision-making is probably overstated, and join Decety's assertion (2021) that empathy is not a reliable and consistent predictor of moral decision-making. However, the lack of a significant predictive effect of cognitive empathy on moral decision-making is still puzzling. As such, while the DPM accurately predicts several outcomes, there appear to be missing pieces to the puzzle that additional features not included in the DPM could likely explain.

We argue that making a decision marks the end of a sequential deliberation process and that moral decision-making is a dynamic process that is impacted by numerous factors. Therefore, both empathy domains may interact at different stages of this process. As a result, some effects may have disappeared, while others may have been reduced by the time the final decision is made (the output).

In conclusion, the key finding of this study was the causal impact of participants' perspectives on moral decision-making. This result highlights the strong influence of this variable on moral dilemmas and underscores the importance of considering this parameter in future research on trolley problems and their variants.

This research investigated several new directions: (1) it confirmed the causal effect of participants' perspective in trolley scenarios and extended this effect by adding a third perspective (Perspective C); (2) it demonstrated that this causal effect also occurred in classical trolley problems (initially, it was only tested on causing inconvenience scenarios); (3) it showed that using everyday trolley-like problems is of great interest as it revealed inter-individual differences and how empathy and BDL traits differently predict moral decision-making; (4) to our knowledge, it is the first study to investigate the role of BDL traits in moral decision-making.

By examining three perspectives on social dilemmas, the trolley problems and variants have become valuable tools for assessing the circumstances under which individuals may prioritize themselves, the community, or the minority. These tools have significant applications in clinical psychology and psychiatry, as they can help clinicians identify specific discrepancies between clinical and healthy control

populations, identify associations with impaired social mechanisms, and offer appropriate treatment and support.

Limitations

There were several limitations in this study. First, similar to previous studies (e.g., Decety & Cowell, 2014; Gleichgerrcht & Young, 2013; Nasello et al., 2021; Patil & Silani, 2014; Takamatsu, 2018), a part of the experimental design was correlational. Thus, while affective empathy and BDL traits significantly predict moral decision-making, they do not necessarily cause it. To overcome this limitation, future research could create conditions that elicit varying degrees of affective empathy and test whether they significantly impact moral decision-making. Second, we used a subclinical measure of borderline traits. Thus, future research can focus on the clinical population to confirm our findings. Third, we did not perform analyses of moral decision-making across genders, as men received men's scenarios and women received women's scenarios. Therefore, we cannot claim these results are generalizable regardless of the target's gender. For instance, future research could use a cross-gender design to investigate whether gender differences obtained in causing harm scenarios are modulated. Finally, the transplant dilemma still faces criticisms similar to that of classical trolley problems raised by Bauman et al. (2014). Due to ethical concerns, only limited experimental approaches can be performed (i.e., using hypothetical scenarios), providing limited evidence. This highlights the need to develop new experimental materials that present more ecologically valid situations for studying moral decision-making in causing harm and causing inconvenience scenarios.

Notes

1. We conducted a pilot experiment ($N = 24$; 17 women; M_{age} = 26.5; $SD = 5.32$) to pretest four everyday trolley-like problems (i.e., the dilemmas of the company, the inheritance, the university, and the roommates) and chose the company's dilemma because: (1) it was rated as as sufficiently immersive and realistic; (2) participants found it highly distressing, and their choices caused them to feel tormented and concerned; (3) participants rated it as presenting low levels of escape (i.e., elements in the scenario that favor one option over the other). Batson et al. (1981) showed that the ease of escape influences helping behaviors in specific empathic conditions. Moreover, (4) participants reported experiencing a wide range of emotions while reading the scenario (i.e., they felt ashamed, anxious, frustrated, disgusted, nervous, and attentive); and (5) they did not find the scenario amusing at all. Taken together, these elements addressed the criticisms raised by Bauman et al. (2014).
2. Another potential explanation for the high percentage of GOD responders among participants with Perspective B could have been the presence of suicidal thoughts, particularly among those with higher levels of BDL traits. However, several factors make this explanation unlikely. First, all participants reported no history of psychiatric or neurological issues. Second, their scores in BDL traits were within the normal range ($Mdn = 115$; $SD = 25$). Third, BDL traits did not significantly predict moral decision-making in Perspective B, ruling out this interpretation.

Acknowledgments

We thank the anonymous reviewers for their helpful comments and suggestions.

Disclosure Statement

The authors report that there are no competing interests to declare.

Fundings Details

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Approval

All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments, and the American Psychological Association's Ethical Principles in the Conduct of Research with Human Participants (2010). The ethical committee of the Department of Psychology of the University of Liège (Belgium) approved the study, reference n°: 1920-92.

Informed Consent

Informed consent was obtained online from all participants included in the study.

Online Data

The online data file is available on the following DOI link: 10.17605/OSF.IO/T4H7F.

Author Notes

Julian A. Nasello is a Ph.D. student at the University of Liège, and his research mainly focuses on the role of empathy in moral decision-making and medical education.

Benoit Dardenne is a social psychology professor at the University of Liège, and his research interests include the role of environment, sexism, and ageism stereotypes in social behaviors.

Michel Hansenne is a professor at the University of Liège in the psychology department, and he specializes in the study of personality and individual differences.

Adélaïde Blavier is a professor at the University of Liège in the psychology department. She has expertise in clinical psychology, especially in psychotraumas, and also works as a forensic psychologist. Her research interests focus on stress and trauma.

Jean-Marc Triffaux is a professor at the University of Liège in the department of medicine. He is a clinician psychiatrist and the medical director of the Psychiatric Day Hospital "La Clé." His research interests mainly focus on the role of empathy in medical care and education.

ORCID

Julian A. Nasello http://orcid.org/0000-0003-0188-1025
Benoit Dardenne http://orcid.org/0000-0002-4001-7101
Michel Hansenne http://orcid.org/0000-0002-3478-4953
Adélaïde Blavier http://orcid.org/0000-0002-2882-4971
Jean-Marc Triffaux http://orcid.org/0000-0002-0058-6190

References

Armstrong, J., Friesdorf, R., & Conway, P. (2019). Clarifying gender differences in moral dilemma judgments: The complementary roles of harm aversion and action aversion. *Social Psychological and Personality Science*, *10*(3), 353–363. https://doi.org/10.1177/1948550618755873

Batson, C. D. (1987). Prosocial motivation: Is it ever truly altruistic? *Advances in Experimental Social Psychology*, *20*, 65–122. https://doi.org/10.1016/S0065-2601(08)60412-8

Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of Personality and Social Psychology*, *40*(2), 290–302. https://doi.org/10.1037/0022-3514.40.2.290

Batson, C. D., Lishner, D. A., & Stocks, E. L. (2015). The empathy—Altruism hypothesis. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford handbook of prosocial behavior* (pp. 259–281). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399813.013.023

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. https://doi.org/10.1111/spc3.12131

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*, *14*(4), 698–718. https://doi.org/10.1016/j.concog.2005.06.004

Bruers, S., & Braeckman, J. (2014). A review and systematization of the trolley problem. *Philosophia*, *42*(2), 251–269. https://doi.org/10.1007/s11406-013-9507-5

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413–425. https://doi.org/10.2307/1128984

Burghart, M., & Mier, D. (2022). No feelings for me, no feelings for you: A meta-analysis on alexithymia and empathy in psychopathy. *Personality and Individual Differences*, *194*, 111658. https://doi.org/10.1016/j.paid.2022.111658

Choe, S. Y., & Min, K. H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making*, *6*(7), 580–592. https://doi.org/10.1017/S193029750000262X

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychological Bulletin*, *135*(3), 495–510. https://doi.org/10.1037/a0015616

D'Ambrosio, F., Olivier, M., Didon, D., & Besche, C. (2009). The basic empathy scale: A French validation of a measure of empathy in youth. *Personality and Individual Differences*, *46*(2), 160–165. https://doi.org/10.1016/j.paid.2008.09.020

Decety, J. (2021). Why empathy is not a reliable source of information in moral decision making. *Current Directions in Psychological Science*, *30*(5), 425–430. https://doi.org/10.1177/09637214211031943

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *9*(5), 525–537. https://doi.org/10.1177/1745691614545130

Decety, J., & Cowell, J. M. (2015). Empathy, justice, and moral behavior. *AJOB Neuroscience*, *6*(3), 3–14. https://doi.org/10.1080/21507740.2015.1047055

DeShong, H. L., Mullins-Sweatt, S. N., Miller, J. D., Widiger, T. A., & Lynam, D. R. (2016). Development of a short form of the five-factor borderline inventory. *Assessment*, *23*(3), 342–352. https://doi.org/10.1177/1073191115581475

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.

Foot, P. (Ed.). (1967). The problem of abortion and the doctrine of the double effect. In *Virtues and vices and other essays in moral philosophy* (pp. 19–32). Basil Blackwell.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality and Social Psychology Bulletin*, *41*(5), 696–713. https://doi.org/10.1177/0146167215575731

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., Lucchiari, C., Consonni, D., Nordio, F., Pravettoni, G., & Cappa, S. (2010a). Gender-related differences in moral judgments. *Cognitive Processing*, *11*(3), 219–226. https://doi.org/10.1007/s10339-009-0335-2

Fumagalli, M., Vergari, M., Pasqualetti, P., Marceglia, S., Mameli, F., Ferrucci, R., Mrakic-Sposta, S., Zago, S., Sartori, G., Pravettoni, G., Barbieri, S., Cappa, S., & Priori, A. (2010b). Brain switches utilitarian behavior: Does gender make the difference? *PLoS One*, *5*(1), e8865. https://doi.org/10.1371/journal.pone.0008865

Gallucci, M. (2019). GAMLj: General analyses for linear models. [Jamovi module]. *Scribbr*. https://gamlj.github.io/.

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PloS One*, *8*(4), e60418. https://doi.org/10.1371/journal.pone.0060418

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154. https://doi.org/10.1016/j.cognition.2007.11.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science (New York, N.Y.)*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Hare, R. D. (1993). *Without conscience: The disturbing world of the psychopaths among us*. Simon & Schuster (Pocket Books).

Hare, R. D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior*, *23*(1), 25–54. https://doi.org/10.1177/0093854896023001004

Hare, R. D., & Neumann, C. S. (2009). Psychopathy: Assessment and forensic implications. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, *54*(12), 791–802. https://doi.org/10.1177/070674370905401202

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language*, *22*(1), 1–21. https://doi.org/10.1111/j.1468-0017.2006.00297.x

Herpertz, S. C., Werth, U., Lukas, G., Qunaibi, M., Schuerkens, A., Kunert, H. J., Freese, R., Flesch, M., Mueller-Isberner, R., Osterheider, M., & Sass, H. (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Archives of General Psychiatry*, *58*(8), 737–745. https://doi.org/10.1001/archpsyc.58.8.737

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*(3), 307–316. https://doi.org/10.1037/h0027580

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas? *PloS One*, *11*(7), e0154780. https://doi.org/10.1371/journal.pone.0154780

Hunt, S. D., & Vitell, S. (1986). A general theory of marketing ethics. *Journal of Macromarketing*, *6*(1), 5–16. https://doi.org/10.1177/027614678600600103

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence*, *29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *The Academy of Management Review*, *16*(2), 366–395. https://doi.org/10.5465/amr.1991.4278958

Kahneman, D. (2013). *Thinking, fast and slow*. Farrar, Straus, and Giroux.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*(1), 160–173. https://www.jstor.org/stable/24966506 https://doi.org/10.1038/scientificamerican0182-160

Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *Lancet (London, England)*, *377*(9759), 74–84. https://doi.org/10.1016/S0140-6736(10)61422-5

Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making*, *27*(1), 57–65. https://doi.org/10.1002/bdm.1782

Nasello, J. A., Blavier, A., & Triffaux, J. M. (2023). French adaptation of the Five-Factor Borderline Inventory-Short Form. *Current Psychology*, 42, 5886–5897. https://doi.org/10.1007/s12144-021-01878-2

Nasello, J. A., Dardenne, B., Blavier, A., & Triffaux, J. M. (2021). Does empathy predict decision-making in everyday trolley-like problems? *Current Psychology*, *42*(4), 2966–2979. https://doi.org/10.1007/s12144-021-01566-1

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem. *Emotion (Washington, D.C.)*, *12*(2), 364–370. https://doi.org/10.1037/a0025561

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*, 501. https://doi.org/10.3389/fpsyg.2014.00501

Peckham, A. D., Jones, P., Snorrason, I., Wessman, I., Beard, C., & Björgvinsson, T. (2020). Age-related differences in borderline personality disorder symptom networks in a transdiagnostic sample. *Journal of Affective Disorders*, *274*, 508–514. https://doi.org/10.1016/j.jad.2020.05.111

Pletti, C., Lotto, L., Buodo, G., & Sarlo, M. (2017). It's immoral, but I'd do it! Psychopathy traits affect decision-making in sacrificial dilemmas and in everyday moral situations. *British Journal of Psychology (London, England : 1953)*, *108*(2), 351–368. https://doi.org/10.1111/bjop.12205

Richetin, J., Preti, E., Costantini, G., & De Panfilis, C. (2017). The centrality of affective instability and identity in Borderline Personality Disorder: Evidence from network analysis. *PloS One*, *12*(10), e0186695. https://doi.org/10.1371/journal.pone.0186695

Robinson, E. V., & Rogers, R. (2015). Empathy faking in psychopathic offenders: The vulnerability of empathy measures. *Journal of Psychopathology and Behavioral Assessment*, *37*(4), 545–552. https://doi.org/10.1007/s10862-015-9479-9

Salgado, R. M., Pedrosa, R., & Bastos-Leite, A. J. (2020). Dysfunction of empathy and related processes in borderline personality disorder: A systematic review. *Harvard Review of Psychiatry*, *28*(4), 238–254. https://doi.org/10.1097/HRP.0000000000000260

Sandoval, A. M. R., Hancock, D., Poythress, N., Edens, J. F., & Lilienfeld, S. (2000). Construct validity of the Psychopathic Personality Inventory in a correctional sample. *Journal of Personality Assessment*, *74*(2), 262–281. https://doi.org/10.1207/S15327752JPA7402_7

Saunders, K. E., Goodwin, G. M., & Rogers, R. D. (2015). Borderline personality disorder, but not euthymic bipolar disorder, is associated with a failure to sustain reciprocal cooperative behaviour: Implications for spectrum models of mood disorders. *Psychological Medicine*, *45*(8), 1591–1600. https://doi.org/10.1017/S0033291714002475

Smillie, L. D., Katic, M., & Laham, S. M. (2021). Personality and moral judgment: Curious consequentialists and polite deontologists. *Journal of Personality*, *89*(3), 549–564. https://doi.org/10.1111/jopy.12598

Smith, K. E., Norman, G. J., & Decety, J. (2017). The complexity of empathy during medical school training: Evidence for positive changes. *Medical Education*, *51*(11), 1146–1159. https://doi.org/10.1111/medu.13398

Southward, M. W., & Cheavens, J. S. (2018). Identifying core deficits in a dimensional model of borderline personality disorder features: A network analysis. *Clinical Psychological Science: A Journal of the Association for Psychological Science*, *6*(5), 685–703. https://doi.org/10.1177/2167702618769560

Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre-and post-decision processes. *Acta Psychologica*, *80*(1-3), 143–168. https://doi.org/10.1016/0001-6918(92)90044-E

Svenson, O. (1996). Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? *Organizational Behavior and Human Decision Processes*, *65*(3), 252–267. https://doi.org/10.1006/obhd.1996.0026

Takamatsu, R. (2018). Turning off the empathy switch: Lower empathic concern for the victim leads to utilitarian choices of action. *PloS One*, *13*(9), e0203826. https://doi.org/10.1371/journal.pone.0203826.

Takamatsu, R. (2019). Personality correlates and utilitarian judgments in the everyday context: Psychopathic traits and differential effects of empathy, social dominance orientation, and dehumanization beliefs. *Personality and Individual Differences*, *146*, 1–8. https://doi.org/10.1016/j.paid.2019.03.029

The Jamovi Project. (2019). Jamovi, version 1.6.23 [Computer Software]. *Scribbr*. https://www.jamovi.org

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204–217. https://www.jstor.org/stable/27902416 https://doi.org/10.5840/monist197659224

Van Dongen, J. D. (2020). The empathic brain of psychopaths: From social science to neuroscience in empathy. *Frontiers in Psychology*, *11*, 695. https://doi.org/10.3389/fpsyg.2020.00695

Vyas, K., Jameel, L., Bellesi, G., Crawford, S., & Channon, S. (2017). Derailing the trolley: Everyday utilitarian judgments in groups high versus low in psychopathic traits or autistic traits. *Psychiatry Research*, *250*, 84–91. https://doi.org/10.1016/j.psychres.2017.01.054

Synthèse de l'étude 2 :

À travers cette étude, nous avons pu démontrer plusieurs points cruciaux. Comme nous l'avions montré dans notre première étude, les choix utilitaires varient considérablement selon la perspective incarnée par le(a) participant(e) dans les différents scénarios. Dans le dilemme de l'entreprise (scénario *causing inconvenience*), on observe que la propension aux choix utilitaires diminue très nettement lorsque le(a) participant(e) doit choisir entre le fait de sauver son emploi ou celui de ses cinq collègues (i.e., les pourcentages sont de 50-50 entre les choix utilitaires et déontologiques), comparativement aux deux autres perspectives où les choix utilitaires dominent largement (> 90%).

En contraste, dans le dilemme du chirurgien (scénario *causing harm*), c'est dans la perspective où le(a) participant(e) se retrouve parmi les cinq patients mourants que les choix orientés vers le groupe sont maximisés (18%). Effectivement, la propension de choix utilitaires double entre cette perspective et celle où le(a) participant(e) incarne le(a) docteur(e) (9%). De plus, on constate qu'il existe une différence significative entre les hommes et les femmes pour ces deux perspectives : les hommes réalisent quasiment deux fois plus de choix utilitaires que les femmes. Toutefois, c'est au niveau de la perspective B (i.e., où le(a) participant(e) incarne le(a) nouveau(elle) patient(e) et où il(elle) doit choisir entre sa vie et celle des cinq autres patients) que les résultats sont étonnants. Dans cette perspective, nous nous attendions à un taux faible de choix utilitaires (le plus faible des trois perspectives). Or, les niveaux de choix utilitaires sont assez comparables à la perspective où le(a) participant(e) incarne le(a) docteur(e).

Dès lors, selon le type de scénario (et les actions qu'ils impliquent) et selon la perspective incarnée par le(a) participant(e), les choix utilitaires vont être modulés de manière probante (et dans certains cas particuliers, selon le genre). Enfin, concernant les traits de personnalité borderline et l'empathie affective, on constate que ces deux variables ont un effet prédicteur « croisé », contrairement à ce qui était attendu. Effectivement, dans le dilemme de l'entreprise (i.e., le scénario *causing inconvenience*), il apparaît que les traits borderlines et l'empathie affective prédisent de manière significative, mais opposée, les choix utilitaires, et ce, uniquement dans la perspective où le(a) participant(e) incarne le(a) représentant(e) : lorsqu'il(elle) doit choisir entre sauver son emploi ou celui de ses cinq collègues. Pour les traits borderlines, plus les traits sont élevés, moins on observera de choix orientés vers le groupe, ce qui rejoint la prédiction faite par le DPM. À l'inverse, pour

l'empathie affective, plus les niveaux sont importants et plus on observera de choix orientés vers le groupe, ce qui ne rejoint pas la prédiction émise à partir du DPM.

Ce dernier résultat rejoint plutôt une hypothèse formulée par Batson (1987 ; 2011), « *the empathy-altruism hypothesis* », montrant que la préoccupation empathique (une dimension centrale de l'empathie affective) génère une motivation altruiste qui se traduit par une plus grande sensibilité à la détresse d'autrui et par des comportements d'aide plus marqués (Batson et al., 2015). On constate donc que les individus présentant des niveaux d'empathie affective plus élevés auront davantage tendance à sacrifier leur emploi pour préserver ceux de leurs collègues.

En ce qui concerne le dilemme du chirurgien (i.e., le scénario *causing harm*), on remarque uniquement un effet prédictif et significatif des traits de personnalité borderline dans la perspective où le(a) participant(e) incarne un des patients mourants : plus les traits sont marqués, plus cela augmente la propension à réaliser un choix orienté vers le groupe (i.e., sacrifier le(a) nouveau(elle) patient(e)). Un résultat qui va également à l'encontre des prédictions émises à partir du DPM.

Que nous enseignent cette deuxième étude ? Premièrement, une fois encore, les effets prédictifs de l'empathie affective et, nouvellement étudiés, des traits de personnalité borderline sur les choix utilitaires sont faibles. Deuxièmement, les prédictions validées émises à partir du DPM sont particulièrement aléatoires. Effectivement, certaines observations sont totalement en opposition avec les prédictions du DPM (comme mentionné ci-avant). Enfin, comme pour notre précédente étude, aucun effet prédictif et significatif n'est observé de l'empathie cognitive sur les choix utilitaires.

Ces différents points dessinent peu à peu un faisceau de preuves démontrant que la théorie du DPM est probablement incomplète ; ce qui rejoint les conclusions de nombreux auteurs (Smillie et al., 2020 ; Oudman et al., 2021 ; Duke & Bègue, 2015 ; Horne & Powell, 2016) qui soulignent cette incomplétude.

Toutefois, nos deux premières études restent focalisées sur des individus issus de la population générale. Il demeure à présent nécessaire d'éprouver cette théorie sur une population clinique d'individus présentant un trouble de la personnalité borderline.

Le choix de cette population clinique part du corpus de données suivantes : premièrement, de nombreuses études ont ciblé l'empathie affective comme caractéristique prédictive des jugements et choix moraux, et la plupart ont montré une association négative entre ces paramètres. Deuxièmement, d'autres études avaient également évalué des populations avec des individus qualifiés comme présentant des traits saillants de psychopathie. Cette population est notamment caractérisée par des scores significativement plus faibles que des participants « contrôles » en empathie affective (Burghart & Mier, 2022), et plusieurs études démontrent qu'ils ont tendance à réaliser plus de choix utilitaires. En étudiant la population clinique à l'opposé du continuum de l'empathie affective, les personnes présentant un trouble de personnalité borderline (Cf., Salgado et al., 2020), nous nous attendons à observer des résultats opposés en termes de prise de décision morale. En effet, il est attendu que ces personnes rapportent significativement moins de choix orientés vers le groupe comparativement à des volontaires dits « sains ».

Chapitre 4 : troisième étude

Dans la continuité des études précédentes, nous avons répliqué le protocole expérimental de l'étude Nasello et al. (2023a ; cf. deuxième étude, Chapitre 3) sur une population clinique de patientes diagnostiquées comme présentant un trouble de la personnalité borderline, en comparaison à des volontaires sains. Dans cette étude, deux nouveaux éléments ont été ajoutés : (1) l'ajout d'un dilemme supplémentaire (le dilemme du départ anticipé : « *The early Departure dilemma* » ; voir Annexe E), un dilemme de type trolley de la vie de tous les jours faisant intervenir une composante de pied-dans-la-porte[6] (Freedman & Fraser, 1966) et (2) une mesure des styles décisionnels. Ces ajouts visaient à déterminer si les patientes présentaient des profils de choix utilitaires différents par rapport aux volontaires sains (il était attendu qu'elles réalisent moins de choix utilitaires) et s'il y avait des éléments propres aux dilemmes moraux qui pouvaient induire une différence significative entre ces populations.

Ensuite, nous cherchions à déterminer si les personnes souffrant d'un trouble grave de la personnalité borderline présentaient des styles de prise de décision spécifiques et si ces styles prédisaient significativement les choix orientés vers le groupe. En effet, les styles de prise de décision sont considérés comme des patterns façonnés par les habitudes et les apprentissages, il est donc probable que ces derniers soient également façonnés par la psychopathologie (ou inversement). De plus, différents auteurs ont montré que les styles de décision ont un impact sur le résultat des décisions qui sont prises par un individu. En d'autres termes, non seulement les individus ont une manière bien à eux de réaliser leurs choix de manière générale, mais cette manière influence le type de choix qu'ils réalisent (Curşeu & Schruijer, 2012). Ceci nous amène à explorer si les styles de décision prédisent significativement les choix moraux.

En résumé, le protocole expérimental comprend trois dilemmes : deux dilemmes de type trolley de la vie de tous les jours (i.e., le dilemme de l'entreprise et le dilemme du départ anticipé) et un dilemme classique (i.e., le dilemme du chirurgien). Comme pour la deuxième étude (cf., Chapitre 3), le participant incarne trois perspectives dans chaque

[6] Le fait qu'un individu accepte une petite requête (généralement peu coûteuse), cela le rend plus susceptible d'accepter une autre requête plus importante. Dans le cadre du dilemme du départ anticipé, le *bystander* n'ayant pas refusé la venue du protagoniste ayant perdu son logement (et le fait qu'il se présente à sa porte) rend le *bystander* plus susceptible de l'accueillir.

dilemme : le « *bystander* », la personne isolée et celui parmi les cinq. Enfin, l'empathie (affective et cognitive) et les styles de prise de décision sont mesurés et nous tenterons de déterminer s'ils prédisent significativement les choix moraux. Le Tableau 1 illustre un récapitulatif et comparatif des études 2 et 3.

Tableau 1 : Comparaison des études 2 et 3.

Etude 2	Etude 3
Echantillon : • Population générale (hommes et femmes) - Individus sans antécédents psychiatriques ou neurologiques **Dilemmes moraux :** • 1 *everyday trolley-like problem* - *The company's dilemma* • 1 dilemme de type trolley « classique » - *The transplant dilemma* • Trois perspectives **Mesures cibles :** • Empathie (BES : Jolliffe & Farrington, 2006) - Empathie affective - Empathie cognitive • Traits borderlines (FFBI-SF : DeShong et al., 2016) **Hypothèses centrales :** • Test de la validité des prédictions émises à partir de la théorie du DPM : - L'empathie affective est un facteur prédicteur négatif des choix utilitaires ; - Les traits borderlines devraient apparaitre comme un facteur prédicteur négatif des choix utilitaires.	**Echantillon :** • Population clinique (uniquement des femmes) - Individus diagnostiqués et volontaires sains **Dilemmes moraux :** • 2 *everyday trolley-like problems* - *The company's dilemma* - *The early departure dilemma* • 1 dilemme de type trolley « classique » - *The transplant dilemma* • Trois perspectives **Mesures cibles :** • Empathie (BES : Jolliffe & Farrington, 2006) - Empathie affective - Empathie cognitive • Traits borderlines (FFBI-SF : DeShong et al., 2016) • Styles décisionnels (Scott & Bruce, 1995) - Rationnel - Intuitif - Dépendant - Evitant - Spontané **Hypothèses centrales :** • Test de la validité des prédictions émises à partir de la théorie du DPM : - L'empathie affective est un facteur prédicteur négatif des choix utilitaires ; - Les traits borderlines devraient apparaitre comme un facteur prédicteur négatif des choix utilitaires. • L'empathie affective est plus élevée chez les individus borderlines comparativement aux volontaires sains. • Les individus borderlines ont des styles décisionnels distincts par rapport aux individus contrôles. • Les styles décisionnels prédisent-ils les choix utilitaires.

Moral decision-making in patients with borderline personality disorder: Are they less inclined towards utilitarianism?

Julian A. Nasello[1,2,*] and Jean-Marc Triffaux[1,3]

1 Psychiatric Day Hospital "La Clé"; Boulevard de la Constitution 153, 4020 Liège (Belgium)

2 Department of Clinical Psychology, University of Liège; Place des Orateurs 1, 4000 Liège (Belgium)

3 Department of Psychiatry, Medicine, University of Liège, CHU of Liège; Avenue de L'Hôptial 1, 4000 Liège (Belgium)

* Corresponding author, email address: julian.nasello@gmail.com

Orcid ID

Julian A. Nasello: 0000-0003-0188-1025

Pr. Jean-Marc Triffaux: 0000-0002-0058-6190

Abstract

Previous studies examined utilitarian choices in trolley dilemmas across various clinical populations (e.g., psychopathy or autism). Building upon Nasello et al.'s (2023) work in the general population, this study focused on patients diagnosed with borderline personality disorder. Both studies aimed to test predictions based on the Dual-Process Model (Greene et al., 2001). It was hypothesized that individuals with borderline traits, characterized by higher affective empathy scores, would exhibit a lower inclination toward utilitarian choices compared to control individuals. In addition, we measured decision-making styles to determine if borderline patients had particular decision-making styles and whether these styles were

associated with utilitarian choices. Forty-three participants voluntarily participated, divided into three groups: patients and two control groups (low and high borderline traits). Three trolley dilemmas were presented, each from three perspectives: bystander, isolated person, and one among five. Results showed that the participants' perspectives and dilemma types significantly influenced moral choices. Borderline traits and affective empathy did not emerge as significant predictors, highlighting the limitations of the Dual-Process Model. Finally, borderline individuals have specific decision-making styles: they appear less rational and more avoidant than control individuals. However, these styles did not significantly predict moral choices. The implications of these findings are discussed.

Keywords

Dual-Process Model; Empathy; Borderline Personality Disorders; Decision-Making; Utilitarianism.

Introduction

A considerable amount of current research is dedicated to understanding moral judgments and decision-making. Various materials, such as the prisoner's dilemma (which explores cooperation) and trolley problems (which investigate utilitarianism), have been created for this purpose.

This paper focuses on moral decision-making within a specific clinical population: individuals diagnosed with borderline personality disorder. It is important to differentiate moral decision-making from moral judgments, as they are often confused (Nasello & Triffaux, 2023). Making a decision marks the end of a sequential deliberation process, where one option is chosen among mutually exclusive alternatives (Nasello et al., 2023a; Nasello & Triffaux, 2023). In contrast, moral judgments arise from an evaluative process that involves assigning positive or negative evaluations to the actions or characteristics of an individual in a given situation

(Haidt, 2001). While moral judgments do contribute to the sequential deliberation process (i.e., the final decision), their influence can vary significantly depending on various factors, such as dispositional traits like psychopathic traits or inherent aspects of the problem. A vast amount of research currently focuses on understanding moral judgments and decision-making. For this purpose, several kinds of materials were created, like the prisoner's dilemma (studying cooperation) and trolley problems (studying utilitarianism).

Would you sacrifice the one to save the five?

If this question seems familiar, it is because it pertains to the well-known trolley problem. The trolley problem, introduced by Foot (1967), presents a scenario where you are the driver of a runaway tram that will hit and kill five workers on the track if no action is taken. Your only available action is to pull a lever and divert the tram to another track where only one worker stands. This dilemma forces you to choose between saving the five and sacrificing the one or saving the one and sacrificing the five. Thomson (1976) later proposed a variant in which the isolated worker needs to be pushed to their death from a bridge to stop the runaway tram. In this variant, the isolated person is obese or carrying a heavy bag and pushing him would stop the tram and save the five workers on the track.

Since then, numerous variations of the trolley problem have emerged in scientific literature, all based on the same principle of favoring one life at the expense of others or vice versa. Authors generally classify a choice to save the five as *utilitarian* and a choice to save the one as *deontological*. These designations align with two major fields in philosophy: utilitarianism, proposed by philosophers such as Bentham (1789), Mill (1861), and Sidgwick (1874), and deontology, developed by Kant (1785). Utilitarian philosophers argue that moral actions should be based on the Greatest Happiness Principle, which means actions are permissible if they maximize overall happiness and minimize pain. Consequences take center stage for utilitarians; hence they are also referred to as consequentialists.

In contrast, deontological philosophers prioritize the moral value of actions and intentions over consequences. In Foot and Thomson's trolley problems, deontologists would choose to save the one because they believe (1) one life is not inherently less important than five and (2) the act of intentionally killing someone, even with the intention to save others, is morally unacceptable.

It is important to note that the classical terminology (Utilitarian vs. Deontological choices) suffers from several limits. In the present paper, we will opt for a more descriptive approach by using the term "group-oriented decisions" instead of "utilitarian decisions." A detailed explanation can be found in Nasello et al. (2023a; 2021).

The Dual-Process Model: A model that might explain moral decision-making?

In the past few decades, authors have attempted to model the mechanisms involved in moral judgments of trolley problems, leading to the development of the Dual-Process Model (DPM, Greene et al., 2001; Greene & Haidt, 2002; Greene et al., 2004; Greene, 2007). The DPM first categorized different types of trolley problems, distinguishing between impersonal dilemmas (such as Foot's trolley problem) and personal dilemmas (similar to Thomson's trolley problem), and then identified the mechanisms at play in these two dilemmas. It was explained that these dilemmas evoke varying levels of cognitive and emotional processes. Foot's problem elicits fewer emotional responses and more cognitive responses, while Thomson's version elicits stronger emotional responses and fewer cognitive responses. Consequently, the involvement of less emotional responses and more cognitive processes increases the likelihood of making utilitarian judgments, whereas the involvement of more emotional responses and fewer cognitive processes increases the likelihood of making deontological judgments. Although this model initially received extensive empirical support, recent research has started to question its significance. The importance attributed to emotional and cognitive processes appears to be weaker than expected, as demonstrated by numerous studies examining the

associations between utilitarian choices, moral judgments, and emotional traits such as empathy, emotions, or alexithymia (Horne & Powell, 2020; Smillie et al., 2021; Nasello et al., 2021; Nasello et al., 2023a; Decety, 2021).

For example, a recent meta-analysis (Nasello & Triffaux, 2023) revealed small to moderate associations between affective empathy (which refers to experiencing the emotions of others; Bryant, 1982) and moral judgments, decision-making, and inclinations. On the other hand, limited or insignificant associations were found between cognitive empathy (i.e., the ability to understand the emotions or mental state of others; Hogan, 1969; Davis, 1983) and moral judgments, decision-making, and inclinations. Horne and Powell (2016) also demonstrated that although moral dilemmas evoke strong emotional responses, they are only weakly associated with moral judgments. Similarly, Smillie et al. (2021) found that only intellect significantly predicted utilitarian inclinations, while other emotion-related traits like neuroticism did not emerge as significant predictors. Overall, these authors have raised doubts about the external validity of the DPM and highlighted its potential incompleteness.

Does psychopathology affect moral decision-making?

Given the substantial variation in proportions of utilitarian choices or judgments depending on the types of trolley problems (impersonal dilemmas ≈ 90% and personal dilemmas ≈ 10%), several authors have investigated whether individuals with psychopathological conditions exhibit different patterns of choices or judgments. For example, studies have examined moral decision-making or judgments in individuals with higher levels of psychopathic traits (Takamatsu, 2018; Takamatsu, 2019; Takamatsu & Takai, 2019; Patil & Silani, 2014; more generally, dark tetrad traits: Dinic et al., 2021), autistic traits (Gleichgerrcht et al., 2013; Patil et al., 2016), borderline traits (Nasello et al., 2023a), anxiety disorders (Patil et al., 2021), or Korsakoff's syndrome (Oudman et al., 2021). Many of these studies have found significant associations between these psychopathological traits or conditions and utilitarian

judgments or choices. For instance, higher levels of psychopathic traits have been associated with a greater proportion of utilitarian choices and judgments (Takamatsu, 2018; Takamatsu, 2019; Takamatsu & Takai, 2019; Patil & Silani, 2014; only in the unknown condition in Dinic et al., 2021). However, findings in other clinical populations have been controversial or non-significant. Studies on individuals with higher levels of autistic traits have yielded contradictory results (Gleichgerrcht et al., 2013; Patil et al., 2016), and no significant differences have been found in Korsakoff's syndrome or anxiety disorders patients (Oudman et al., 2021; Patil et al., 2021).

Regarding borderline personality traits, one study revealed different significant associations between these traits and moral decision-making depending on the type of trolley problem and the participant's perspective. Borderline personality disorder (BPD) is a severe and persistent psychiatric disorder characterized by emotional instability, unstable interpersonal relationships, disrupted self-image, and impulsive behavior. While BPD shares some symptoms with psychopathy (e.g., impulsivity), it significantly differs in terms of emotional reactivity. Individuals with BPD exhibit heightened emotional reactivity (Crowell et al., 2009), whereas psychopaths display emotional callousness (Hare, 1996; Hare & Neumann, 2009) and reduced responsiveness (Herpertz et al., 2001). In terms of empathy, these two psychopathological populations also differ significantly in affective empathy but show similar lower levels of cognitive empathy (Salgado et al., 2020; Burghart & Mier, 2022). Specifically, individuals with BPD have higher levels of affective empathy (Salgado et al., 2020), while psychopaths have lower levels (Burghart & Mier, 2022).

In summary, consistent findings have shown that individuals with higher levels of psychopathic traits are more inclined to make group-oriented choices or judgments. This supports the DPM as these individuals are less emotionally responsive to moral dilemmas, leading to greater involvement of cognitive processes in moral decision-making. However, the

conclusions are less clear for individuals who are more emotionally aroused. According to the DPM, they would be expected to make fewer group-oriented choices or judgments. To address this prediction of the DPM, we aim to conduct a study in a female clinical population with borderline personality disorder (BPD). This study builds upon previous research conducted in the general population (Nasello et al., 2023a), where we assessed borderline personality traits to examine whether they predict lower levels of group-oriented choices. We utilized two types of trolley problems, namely an everyday trolley-like problem (the company's dilemma) and a classical personal trolley problem (the transplant dilemma). All participants considered three perspectives in these two trolley-type problems: the bystander (P_A), the isolated person (P_B), and the one among the five (P_C). Our previous study demonstrated significant and distinct predictive effects of borderline personality traits in the two trolley-type problems. Furthermore, these effects were significant, albeit with small effect sizes, in specific perspectives. In the company's dilemma, borderline traits significantly decreased the likelihood of making a group-oriented choice in P_B (when individuals had to choose between sacrificing their own job or the jobs of their five colleagues). Conversely, these traits significantly increased the likelihood of making a group-oriented choice in P_C of the transplant dilemma (where higher levels of borderline personality traits increased the propensity to sacrifice an isolated organ-compatible patient to save the five dying patients, but only when the participant was one among the five patients). Therefore, the present study aims to replicate these findings in a BPD population and compare them to healthy control participants.

Objectives and hypotheses

The main objective of the present study is to replicate the findings of Nasello et al. (2023a) in a clinical population of individuals with borderline personality disorder (BPD). Using a general population sample, Nasello et al. (2023a) demonstrated that the participant's perspective plays a causal role in moral decision-making, with variations depending on the type

of dilemma (everyday trolley-like problem vs. classical trolley problem). They also found significant associations between borderline traits and moral decision-making, with a negative association in P_B of the company's dilemma and a positive association in P_C of the transplant dilemma (as discussed earlier). Additionally, they found that affective empathy had a significant positive predictive effect on group-oriented choices only in P_B of the company's dilemma. We anticipate similar findings in the present study. A new component has been introduced by including an additional everyday trolley-like problem (the early departure: a dilemma involving a foot-in-the-door feature, see Materials section). Furthermore, we measured the decision-making styles (rational, avoidant, intuitive, spontaneous, and dependent) of participants to investigate (1) potential differences in decision-making styles between BPD patients and healthy controls and (2) whether these decision-making styles significantly predict group-oriented choices.

Individuals are known to exhibit significant variations in their decision-making approaches (Girard et al., 2016). Decision-making styles, which represent patterns of decision-making shaped by learning and habits, have been identified, with five styles proposed by Scott and Bruce (1995): rational, intuitive, avoidant, spontaneous, and dependent. Previous research has shown that decision styles can predict decision behaviors (Curşeu & Schruijer, 2012); thus, we anticipate a significant association between decision-making styles and group-oriented choices.

Method

Participants and design

The clinical sample comprised 15 female outpatients diagnosed with BPD (M_{age} = 31.3; *SD* = 8.52): all BPD patients were Caucasian females (see Table 1 for other sociodemographic information). These participants were recruited through advertising distributed to psychiatrists

in the Liège locality (Belgium). Inclusion criteria required patients to be 18 years or older, diagnosed with BPD and aware of this diagnosis, have a minimum four-week period of abstinence from problematic substance abuse, not have severe organic illness, mental retardation, or a major neurological history, and not have a history of schizophrenia or other psychotic disorders (information confirmed by their psychiatrist). Psychiatrists provided a list of potential patients meeting the study criteria and asked their patients if they would be willing to participate. Upon agreement, the researcher contacted the patients, provided detailed information about the study, and scheduled a meeting appointment. During the meeting, the experimental session started after the participant provided informed consent and lasted approximately 90 minutes. The session began with a structured clinical interview for DSM-V (APA, 2013) Axis II personality disorders using the SCID-II (First et al., 1994) to determine if the BPD patient met the BPD criteria. The median number of BPD criteria met was 8 (*Min*. = 6; *Max*. = 9). Subsequently, the participant completed several questionnaires, including sociodemographic data collection, two scales measuring borderline personality traits, an empathy scale, an anxiety and depression scale, and a decision-making styles questionnaire (refer to the Materials section for a comprehensive description). Finally, the participant was presented with three randomly distributed trolley-type problems (the company's dilemma, the early departure dilemma, and the transplant dilemma) and was asked to choose between the two options for each problem across three perspectives (P_A: the bystander, P_B: the isolated person, and P_C: the one among the five). One patient was not included because she scored 4 on the SCID-2 and had comparable borderline trait scores to both control groups.

The healthy control (HC) group was composed of 15 females (M_{age} = 31.2; *SD* = 9.37). They were recruited through social media and bulletin-board announcements. Each control woman was matched with each BPD patient regarding age, education, and socio-economic status (see Table 1). HC participants should not present any problematic substance abuse, severe organic illness, mental retardation, or neurological history. The procedure was the same

for BPD patients and HC, except that HC did not pass the semi-structural interview (SCID-2, First et al., 1994).

An unexpected third group was composed during the study. Indeed, 50% of HC presented higher-than-average levels of borderline personality traits during the recruitment phase. An arbitrary cut-off score of ≥125 at the FFBI-SF was determined to distinguish between HC with low and high levels of borderline traits (or 142 and more when using the total borderline trait score). This sample, like HC, did not present any problematic substance abuse, severe organic illness, mental retardation, or psychiatric or neurological history. The 13 females (M_{age} = 30.6; *SD* = 9.46) in this group scored significantly higher than HC but significantly lower than BPD patients in the two borderline personality traits scales (see Table 2). As they were initially HC participants, they matched with BPD patients regarding age, education, and socio-economic status (see Table 1). One participant was not included because she reported a history of psychiatric issues (depression) and disclosed undergoing antidepressant treatment.

Therefore, the present study consists of three groups: BPD patients (*n* = 15), HC with low levels of borderline traits (HC_L: *n* = 15), and HC with high levels of borderline traits (HC_H: *n* = 13).

Table 1. Sociodemographic information

	BDL patients (n = 15)	HC_H (n = 13)	HC_L (n = 15)	χ^2
Sex	Female (100%)	Female (100%)	Female (100%)	-
Ethnicity	Caucasian (100%)	Caucasian (100%)	Caucasian (100%)	-
Marital status	Single (60%) Legal cohabitant (13%) Married (0%) Divorced (20%)	Single (69%) Legal cohabitant (31%) Married (0%) Divorced (0%)	Single (67%) Legal cohabitant (20%) Married (0%) Divorced (13%)	p = .472
Education	Master's degree (7%) Bachelor's degree (40%) ≤ High school degree (53%)	Master's degree (8%) Bachelor's degree (54%) ≤ High school degree (38%)	Master's degree (7%) Bachelor's degree (33%) ≤ High school degree (60%)	p = .601
Socioeconomic status*	High (13%) Higher than average (20%) Average (7%) Lower than average (13%) Low (47%)	High (15%) Higher than average (20%) Average (8%) Lower than average (15%) Low (38%)	High (7%) Higher than average (20%) Average (0%) Lower than average (33%) Low (40%)	p = .901

*__Socioeconomic status__: High = > 1800 euros net per month; Higher than average = between 1600 and 1800 euros net per month; Average = between 1600 and 1800 euros net per month; Lower than average = between 1200 and 1400 euros net per month; Low = < 1200 euros net per month.

Materials

Sociodemographic data collection. We collected information such as age, socio-economic status, marital status, current main occupation, highest attained degree, medication(s), medical, psychiatric, and neurological history, and finally, the history of past substance use (subjective importance attributed to these substances and frequency of use).

Empathy. The Basic Empathy Scale (Jolliffe & Farrington, 2006; French version: D'Ambrosio et al., 2009) is a 20-item questionnaire that measures two aspects of empathy (affective empathy and cognitive empathy) and focuses on four basic emotions (i.e., anger, fear, happiness, and sadness). Affective empathy refers to the ability to share and experience the emotions of others (Bryant, 1982), and it comprises 11 items. On the other hand, cognitive

empathy is the capacity to understand the emotions of others (Hogan, 1969), and it includes nine items. Participants rate each item on a 5-point Likert scale ranging from 1 to 5.

Borderline Traits. To assess subclinical borderline traits, we utilized the Five-Factor Borderline Inventory, Short Form (FFBI-SF, DeShong et al., 2016; French version: Nasello et al., 2023b), a self-rated questionnaire consisting of 48 items. The scale consists of assessing 12 domains (anxious uncertainty, dysregulated anger, despondence, self-disturbance, oppositional, behavior dysregulation, affective dysregulation, fragility, dissociative tendencies, distrustfulness, manipulativeness, and rashness). Participants were asked to rate each item on a 5-point Likert scale, ranging from 1 to 5. A total score can be calculated, and individuals scoring ≥162 on this scale could be considered as having salient borderline traits (Nasello et al., 2023b). As a recent scale measuring borderline personality traits, we also used the Borderline Evaluation of Severity over Time (BEST; Pfohl et al., 2009). This 15-item scale measures three domains: thoughts and feelings (rating mood reactivity, identity disturbance, unstable relationships, paranoia, emptiness, and suicidal thinking), negative behaviors (rating negative actions like mutilations), and positive behaviors (rating newly learned adaptive skills). Items from the two first domains are rated from 1 to 5 and from 5 to 1 for the later domain.

The BEST and FFBI-SF displayed good psychometric properties (Pfhol et al., 2009; DeShong et al., 2016; Nasello et al., 2023b), especially good internal consistency.

Decision-making styles. Decision-making styles were assessed using the General Decision-Making Style Questionnaire (Scott & Bruce, 1995; French version: Girard et al., 2016). This questionnaire measured five domains: *rational* (the logical seeking and assessment of alternative choices), *intuitive* (relying on gut feelings and instincts for decision-making), *dependent* (seeking guidance and recommendations from others to make decisions), *avoidant* (making efforts to avoid decisions), and *spontaneous* (efforts in making decisions as quickly as possible). The questionnaire comprises 25 items (5 for each style) rated on a 5-point Likert

scale. The scale demonstrates adequate psychometric properties, including acceptable internal consistency (Scott & Bruce, 1995; Girard et al., 2016).

Anxiety and Depression. We used the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983; French version: Bocéréan & Dupret, 2014) to measure the anxiety and depression levels of our participants. This scale comprises 14 items (7 for depression and 7 for anxiety) coded from 0 to 3 and displayed good psychometric properties (see Zigmond & Snaith, 1983; Bocéréan & Dupret, 2014), notably good internal consistency.

All Cronbach's alphas are reported in Table A in the Supplementary materials.

Trolley-like problems. Three trolley-type problems (one classical trolley problem and two everyday trolley-like problems), each with three perspectives (i.e., P_A: the bystander, P_B: the isolated person, and P_C: the one among the five), were used in this study (see Tables B & C in the Supplementary materials).

We used a classical trolley problem called the transplant dilemma. In one perspective of this scenario, participants took on the role of a skilled surgeon facing a difficult decision (P_A). Five patients are critically ill due to organ failure, with no chance of survival without a transplant. Meanwhile, a new patient arrives with a non-life-threatening condition, and the surgeon realizes that the organs of this new patient are compatible with those of the five dying patients. The surgeon has two options: (1) perform a lethal maneuver during the new patient's surgery, sacrificing the new patient's life to save the five dying patients, or (2) proceed with the new patient's surgery following standard protocols, ensuring her survival but resulting in the death of the other five patients. Participants were randomly assigned roles as the surgeon (P_A), the new patient (P_B), or one of the five dying patients (P_C).

The second dilemma, known as the company's dilemma, involves a situation where a small company is facing the need to downsize its workforce due to the ongoing pandemic. The

participants in this scenario are employees of the company, including the boss (P_A), the spokesperson (P_B), or one of the five workers (P_C). The employees, worried about their job security and termination compensation, decide to go on strike. To address the situation, the boss arranges a meeting with the spokesperson to discuss the matter. During the meeting, the boss explains the limited options available and proposes a two-option arrangement behind closed doors without involving the other employees. The first option is to protect the spokesperson's job while laying off the other five employees without termination compensation. The second option is to lay off the spokesperson without compensation but protect the jobs of the other five employees.

The third trolley problem, known as the early departure dilemma, involves a situation where a friend (e.g., Julie) urgently needs a place to stay as she has to leave her current apartment two weeks before moving into her new one. A friend lives with five other friends who are currently under a lot of stress due to final exams. These five roommates do not get along with Julie and have had arguments with her in the past. Unable to answer the call, you receive a voicemail from Julie explaining her situation. She plans to come and stay at your place starting tonight. Later in the evening, Julie shows up at your door, and without consulting your roommates, you must choose between allowing Julie to stay for the next two weeks (and the four other roommates will have to accommodate her during their exam period) or not. The participants in this scenario randomly enrolled the friend (P_A), Julie (P_B), or one of the five roommates (P_C), and this scenario includes a foot-in-the-door feature (once individuals comply with a small request, they are more likely to comply with a larger demand; Freedman & Fraser, 1966).

It is important to note that the decision-maker in each scenario makes the decision individually without involving the other parties. The scenarios were designed with same-gender scenarios presented to participants in Perspectives A and C.

Data Analyses and Experimental Design

The main analysis utilized a Generalized Mixed Model (GMM) to examine the predictive factors of participants' perspectives, affective empathy, and BDL traits on moral decision-making in trolley problems and their variations. We did not include the decision-making style variable because no simple regression analysis demonstrated a significant predictor effect of any decision-making style domain on group-oriented decisions, see Table I in the Supplementary materials). Three separate GMMs were run for each trolley-type scenario. In these analyses, group-oriented choices were coded as "1" and other choices as "0". Random effects included intercept and ID, while all predictors were centered. A bootstrap was applied in the CI method. For a comprehensive explanation of the GMM procedure, refer to Gallucci (2019).

Descriptive statistics involved chi-square tests to identify choice differences between the three types of scenarios across the three perspectives, Spearman's correlations were performed to study the links between specific variables, and Friedman's ANOVAs for paired samples were also conducted to identify significant group differences in all measurements (Kendall's W were used to determine the effect size). The analyses were conducted using JAMOVI computer software, version 1.6.23 (The Jamovi Project, 2019).

Regarding sample size, there is no specific software that allows for the determination of the sample size for a GMM. However, G*Power (version 3.9.1.7; Faul et al., 2007) can be used to estimate the required sample size for conducting a logistic regression analysis, which is based on the same principles as a GMM. With an odds ratio of 1.55, a significance level (α) of 0.05 (two tails), a power of 0.80, and 3 predictor parameters, G*Power suggested that a sample size of 41 was necessary, meeting other authors' requirement of 10 or more events per variable in logistic regressions (e.g., Peduzzi et al., 1996). Note that the study was not preregistered, but all measures, manipulations, and exclusions have been reported.

Results

Descriptive statistics

Regarding the measurement of borderline traits, we identified three distinct groups (the three groups strongly differed, see Table 2). With the exception of the *positive behaviors* domain of the BEST, all groups showed significant differences in the two other domains (*thoughts and feelings* and *negative behaviors*), with the BPD group displaying significantly higher scores than HC_L and HC_H. In terms of the FFBI-SF, HC_L had significantly lower scores than BPD across all domains (see Table 2 & Tables D-E in the Supplementary materials). However, HC_L and HC_H exhibited similar levels of anxious uncertainty and rashness but differed significantly in all other FFBI-SF domains. HC_H and BPD patients demonstrated similar levels of anxious uncertainty, dissociative tendencies, distrustfulness, and manipulativeness (but differed significantly in other domains). When significant, a gradient in scores was observed (HC_L < HC_H < BPD group).

In terms of depression and anxiety, we found moderate to strong differences between the groups: HC_L had significantly lower scores than HC_H and BPD, while HC_H and BPD exhibited similar levels (see Table 2 & Tables D-E in the Supplementary materials). Regarding affective and cognitive empathy, no significant differences were found among the three groups ($ps > .05$; see Tables F-G in the Supplementary materials), although some marginal correlations were observed between affective empathy and some measurements of borderline traits ($rs \geq .25$; $ps < .10$; see Table H in Supplementary Materials).

Finally, HC_H and BPD had significantly higher scores than HC_L (small effect) in terms of the avoidant style of decision-making (no difference was found between HC_H and BPD), and BPD had significantly lower scores than HC_L and HC_H (moderate effect) in the rational decision-making style (see Table 2 and F-G in the Supplementary materials).

Moral decision-making differences

The company's dilemma. Significant differences were observed between perspectives ($\chi^2_{(2)}$ = 27.7; p < .001). Specifically, P_B differed significantly from both P_A and P_C (ps < .001), while P_A and P_C exhibited similar levels of group-oriented choices (p = .112). **The early departure dilemma**. Participants' choices varied significantly depending on the perspectives ($\chi^2_{(2)}$ = 20.7; p < .001): P_B significantly differed from P_A and P_C (ps < .001), while P_A and P_C showed similar levels of group-oriented choices (p = .168). **The transplant dilemma.** No significant choice differences were found between the three perspectives ($\chi^2_{(2)}$ = 2.00; p = .368). Figure 1 depicts choice differences.

Across all trolley-type problems and perspectives, no significant differences were found among the three groups (ps > .65), indicating that they made similar levels of group-oriented choices.

Moral decision-making and decision-making styles

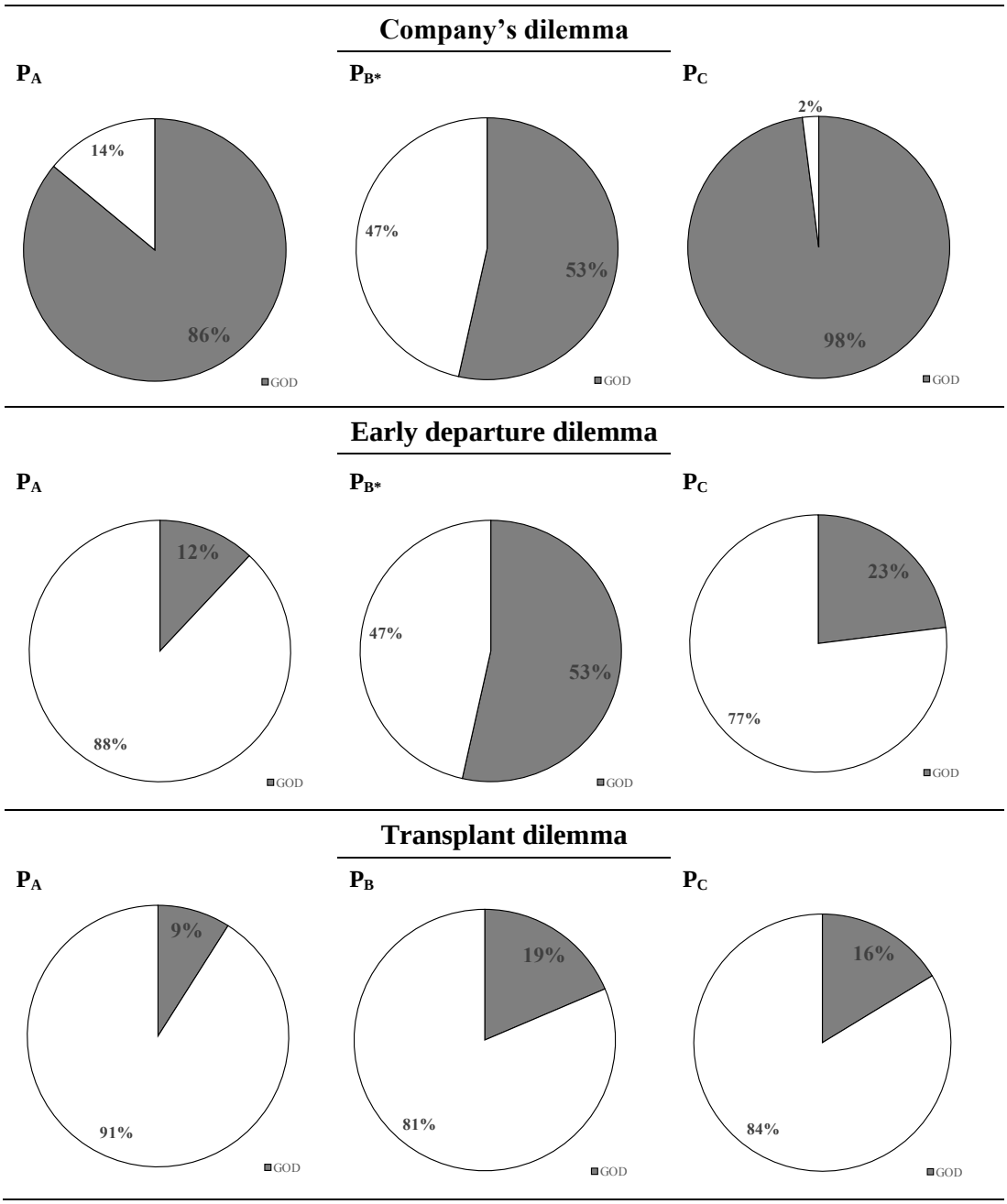
None of the decision-making style domains emerged as a significant predictor of group-oriented choices. However, significant correlations were observed between these domains and certain perspectives of trolley scenario types (refer to Tables I & J in the Supplementary materials). Two positive correlations were found: one between the spontaneous style and P_A of the company's dilemma (r = .32; p < .05) and another between the avoidant style and P_A of the early departure dilemma (r = .31; p < .05). Additionally, a negative correlation was discovered between the dependent style and P_C of the transplant dilemma (r = -.30; p < .05).

Table 2. Group differences in borderline traits, anxiety and depression, empathy, and decision-making styles.

	BPD (***n*** **= 15)**	**HC_H** (***n*** **= 13)**	**HC_L** (***n*** **= 15)**			
	***M* (*SD*)**	***M* (*SD*)**	***M* (*SD*)**	$\chi^2_{(2)}$	***p***	***W***
Age	31.3 (8.52)	30.6 (9.46)	31.2 (9.4)	.776	.679	-
Borderline traits						
BEST	34.3 (6.21)	23.1 (5.42)	10.5 (7.04)	21.7	**< .001**	**.723**
FFBI-SF	178.3 (17.7)	146.5 (15.8)	93.8 (20.9)	26.0	**< .001**	**.866**
Total borderline scores	212.7 (22.8)	169.6 (19.7)	104.3 (24.5)	25.5	**< .001**	**.850**
Anxiety and depression						
Anxiety	14.7 (3.94)	13.2 (3.66)	8 (3.61)	10.7	**.005**	**.357**
Depression	8.93 (3.92)	7.62 (3.69)	3.33 (2.02)	13.0	**< .001**	**.433**
Total A. & D.	23.6 (6.80)	20.9 (5.58)	11.3 (4.79)	16.6	**< .001**	**.553**
Empathy						
Affective empathy	45.9 (7.85)	45.5 (4.41)	41.9 (7.85)	1.92	.383	-
Cognitive empathy	38.6 (5.12)	38.4 (5.08)	38.2 (5.52)	.163	.922	-
Decision-making styles						
Intuitive	17.9 (3.89)	17.4 (3.28)	18.9 (2.58)	1.54	.463	-
Dependent	21.4 (2.50)	19.5 (2.79)	19.2 (3.67)	4.27	.118	-
Avoidant	19.5 (3.80)	20.5 (3.43)	13.5 (3.72)	6.90	**.032**	**.230**
Spontaneous	17.6 (3.72)	14.9 (3.78)	13.5 (5.82)	2.68	.262	-
Rational	17 (4.28)	18.2 (3.26)	19.6 (1.77)	9.04	**.011**	**.301**
Group-oriented decisions	***N* (GOD%)**	***N* (GOD %)**	***N* (GOD %)**	$\chi^2_{(2)}$	***p***	
Company's dilemma	36 (80%)	32 (82.1%)	34 (75.6%)	.639	.727	
P_A	14 (93.3%)	11 (84.6%)	12 (80%)	.378	.828	
P_B	7 (46.7%)	8 (61.5%)	8 (53.3%)	.087	.957	
P_C	15 (100%)	13 (100%)	14 (93.3%)	.143	.931	
Early departure dilemma	14 (31.1%)	15 (38.5%)	9 (20%)	4.74	.093	
P_A	2 (13.3%)	2 (15.4%)	1 (6.7%)	.400	.819	
P_B	8 (53.3%)	9 (69.2%)	6 (40%)	.609	.738	
P_C	4 (26.7%)	4 (30.8%)	2 (13.3%)	.800	.670	
Transplant dilemma	7 (15.6%)	5 (12.8%)	7 (15.6%)	.644	.725	
P_A	2 (13.3%)	1 (7.7%)	1 (6.7%)	.500	.779	
P_B	3 (20%)	2 (15.4%)	3 (20%)	.250	.882	
P_C	2 (13.3%)	2 (15.4%)	3 (20%)	.286	.867	
Total GOD	57 (42.2%)	52 (44.4%)	50 (37%)	2.16	.339	

χ^2 and Friedman's analyses. *Abbreviations*: BEST = Borderline Evaluation of Severity over Time; BPD = Borderline Personality Disorders group; FFBI-SF = Five-Factor Borderline Inventory, Short Form; GOD = Group-Oriented Decisions; HC_H = Healthy Controls displaying high levels of borderline traits; HC_L = Healthy Controls displaying low levels of borderline traits; $P_{A,B,C}$ = Perspective A (the bystander), B (the isolated person), or C (the one among the five); Total A. & D. = Total Anxiety and Depression.

Figure 1. Moral decision-making in the three trolley-type dilemmas.



Abbreviations: GOD = Group-Oriented Decisions (utilitarian choices); $P_{A,B,C}$ = Perspective A (the bystander), B (the isolated person), or C (the one among the five). *: $p = .761$

Generalized Mixed Model analyses

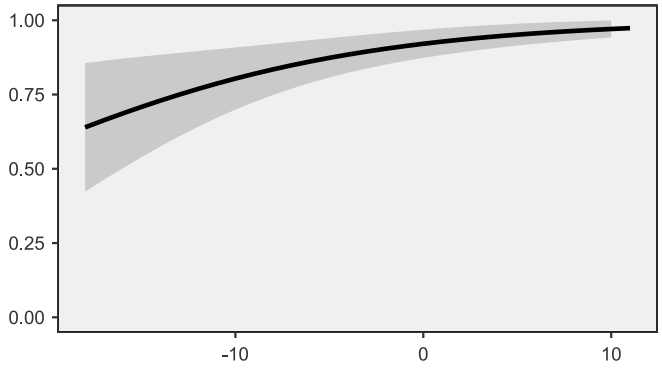
The company's dilemma. The GMM ($\chi 2/df$ = 0.501; $R^2_{Marg.}$ = .452; $R^2_{Cond.}$ = .628) revealed that the participant's perspective strongly predicted group-oriented choices ($\chi^2_{(2)}$ =

13.1; p = .001; $P_{A\text{-}B}$: B = -2.28; *exp(B)* = 0.102; p < .001; $P_{A\text{-}C}$: B = 2.23; *exp(B)* = 9.28; p = .067; see Table 3). Affective empathy marginally predicted group-oriented choices (B = .105; *exp(B)* = 1.110; p = .055; see Table 3 and Figure 2), but no significant predictive effect was found regarding borderline traits (B = .002; p = .781). **The early departure dilemma**. The GMM ($\chi 2/df$ = 0.694; $R^2_{Marg.}$ = .263; $R^2_{Cond.}$ = .438) showed a strong predictive effect of the participant's perspective ($\chi^2_{(2)}$ = 15.2; p < .001; $P_{A\text{-}B}$: B = 2.67; *exp(B)* = 14.4; p < .001; $P_{A\text{-}C}$: B = 0.972; *exp(B)* = 2.64; p = .135; see Table 3). However, neither affective empathy (B = -.05; p = .204) nor borderline traits (B = .010; p = .110) significantly predicted group-oriented choices. **The transplant dilemma.** No significant predictive effects of perspective (ps > .15), borderline traits (p = .828), or affective empathy (p = .159; see Table 3) were found in the GMM analysis ($\chi 2/df$ = 0.509; $R^2_{Marg.}$ = .081; $R^2_{Cond.}$ = 0.438).

A Spearman correlation matrix accompanies the GMM analyses (see Table K in the Supplementary materials).

Figure 2. Effect plots of affective empathy on group-oriented decisions

Company's dilemma



Note. The X-axis represents the affective empathy variable, while the Y-axis indicates the degree of group-oriented choices, with values closer to 1 indicating a stronger inclination towards such choices. The standard error is displayed in gray.

Table 3. Coefficients of the General Mixed Model Analysis.

					95% exp(B) CI			
	Variables	**Estimate**	**SE**	**exp(B)**	**Lower**	**Upper**	**z**	**p**
Company dilemma	(Intercept)	2.46	0.652	11.6	6.81	2.1^{e+10}	3.77	< .001
	P_A-P_B	-2.28	0.775	0.102	2.7^{e-7}	0.405	-2.95	0.003
	P_A-P_C	2.23	1.22	9.28	1.673	2^{e+7}	1.83	0.067
	BDL traits	0.002	0.007	1.002	0.985	1.172	0.278	0.781
	AE	0.105	0.054	1.110	1.000	5.09	1.92	0.055
Early departure dilemma	(Intercept)	-1.28	0.341	0.279	0.104	0.473	-3.74	< .001
	P_A-P_B	2.67	0.715	14.4	4.88	91.7	3.73	< .001
	P_A-P_C	0.972	0.651	2.64	0.787	14.8	1.49	0.135
	BDL traits	0.010	0.006	1.01	0.999	1.027	1.60	0.110
	AE	-0.053	0.042	0.948	0.859	1.025	-1.27	0.204
Transplant dilemma	(Intercept)	-2.52	0.616	0.081	0.024	0.270	-4.08	< .001
	P_A-P_B	1.006	0.746	2.73	0.634	11.8	1.35	0.177
	P_A-P_C	0.796	0.752	2.22	0.508	9.68	1.06	0.290
	BDL traits	0.002	0.008	1.002	0.987	1.017	0.217	0.828
	AE	-0.078	0.056	0.925	0.830	1.031	-1.41	0.159

Note. A positive estimate reveals an increase in the likelihood of making a group-oriented choice. *Abbreviations*: AE = Affective empathy; BDL Traits = Borderline Traits; $P_{A,B,C}$ = Perspective A (the bystander), B (the isolated person), or C (the one among the five).

Discussion

Borderline personality disorders and personality characteristics

Regarding measurements of borderline traits, female BPD patients exhibited higher levels of negative behaviors, dysregulated anger, despondence, self-disturbance, behavioral and affective dysregulation, fragility, oppositionality, and rashness compared to HC_L and HC_H, showing that these traits are particularly specific to borderline personality disorder. However, similar levels of positive behaviors were observed across the three groups. BPD patients also reported higher levels of depression and anxiety compared to HC_L, while similar levels of anxiety and depression were found between BPD and HC_H. This suggests that people presenting high levels of borderline traits may be more vulnerable to depression and anxiety.

There were no significant differences in empathy domains between BPD, HC_H, and HC_L. However, some marginal positive correlations were noted between affective empathy and some borderline traits, which were partially in line with the findings of Salgado et al. (2020).

Regarding decision-making styles, BPD patients displayed higher levels of avoidance and lower levels of rational styles compared to HC_L: similar scores were observed between HC_H and BPD in the avoidance style and between HC_H and HC_L in the rational style. These particular styles of decision-making in BPD may be closely associated with avoidant behaviors, such as avoiding abandonment or intense emotional experiences (Sauer-Zavala & Barlow, 2014; see also McCloskey et al., 2020), as well as impulsive behaviors, like suicide attempts, self-mutilation, and risky behaviors. Although none of the decision-making styles significantly predicted moral decision-making, there were significant correlations found between the dependent, avoidant, and spontaneous styles and specific perspectives of the trolley-type problems. Specifically, higher levels of spontaneous and avoidant styles were associated with increases in group-oriented decisions, while higher levels of dependent style were associated with decreases in group-oriented decisions. It is important to exercise caution when interpreting these correlations, considering the small sample size of our study. Nevertheless, these findings suggest a potential new field of investigation for future research.

Borderline personality traits and affective empathy in trolley-type problems

Our study found that female BPD patients demonstrated similar levels of group-oriented choices compared to both healthy control groups. This finding contradicts the predictions of the DPM (Greene et al., 2001; 2004), as we would have expected lower group-oriented choices in the BPD group. Interestingly, the trend appears to be in the opposite direction, with higher overall group-oriented choices found in the BPD and HC_H groups, although these differences were not statistically significant. Contrary to Nasello et al.'s findings (2023a), our study using GMMs did not reveal any significant predictive effects of borderline

traits on group-oriented choices. Nasello et al. (2023a) reported that borderline traits significantly predicted group-oriented choices in P_B of the company's dilemma (when participants had to choose between their own job or their five coworkers' jobs) and P_C of the transplant dilemmas (when participants were one among the five dying patients). However, in a smaller sample containing BPD patients, these findings were not replicated.

The DPM posits that heightened emotional experiences override cognitive processes, resulting in a decrease in group-oriented decisions. According to this theory, we posited that higher levels of affective empathy should further amplify the influence of emotional processes, leading to an even greater reduction in group-oriented decisions. GMMs indicated only marginal predictive effects of affective empathy on group-oriented choices in the company's dilemma. Interestingly, a marginal association was observed between higher levels of affective empathy and a greater tendency to make group-oriented choices in scenarios causing major inconvenience (in the company's dilemma), which aligns with Nasello et al.'s findings (2023a) but contradicts the DPM theory (Greene et al., 2001; 2004). However, it is worth noting that this affective empathy effect was only marginal. This result supports authors who argued that empathy is not the most relevant factor in predicting moral choices (Decety, 2021; Decety & Cowell, 2014, 2015). More broadly, it is likely that empathy is frequently and erroneously used to explain certain unexpected or even unexplained moral behaviors (as it is wrongly considered essential in moral conduct).

Perspectives and Trolley-Type Problems: Influences on group-oriented choices

The GMMs showed some strong predictive effects of the participant's perspective in everyday trolley-like problems, such as the company and early departure dilemmas, while an insignificant predictive effect was observed in the transplant dilemma. In both everyday trolley-like problems, perspective B had the most significant impact on participants' decision-making. For example, in the early departure dilemma, shifting from perspective A to B resulted in a 14-

fold increase in the odds of making a group-oriented choice. In contrast, in the company's dilemma, the odds of making a group-oriented choice decreased by 90% when participants transitioned from perspective A to B (and moving from perspective A to C only marginally increased the odds of making a group-oriented choice in the company's dilemma). These results are consistent with previous studies (Nasello et al., 2021; 2023a) and provide a body of evidence demonstrating that the participant's perspective in a moral dilemma has a strong causal effect on group-oriented decision-making, highlighting the need to consider this variable when studying moral decision-making processes. Moreover, the participant's perspective effect is not significantly affected by borderline personality disorders (at least for females).

Another factor playing an important modulation of group-oriented choices is the trolley-type problem. This factor has been known to play a significant role since the beginning of experiments using these scenarios. Remember, the classification between Foot's (impersonal) scenario and Thomson's (personal) scenario already led to drastically opposing moral choices (respectively, around 90% vs. 10% of group-oriented decisions in the bystander perspective). In the present study, we clearly observe that scenarios involving causing major inconvenience (the company's dilemma), compared to a problem involving the foot-in-the-door phenomenon (the early departure dilemma), or even intentionally causing death (the transplant dilemma), will significantly modulate decision-making patterns. Once again, this factor was not significantly affected by borderline personality disorders.

Considering their effect sizes, the main conclusion drawn from these effects (i.e., the participant's perspective and types of trolley problems) on group-oriented decisions is that the moral context of these scenarios and the endorsed perspective have a more significant influence on individuals' moral decision-making than their personality traits. In other words, we propose that the moral context in which participants are immersed, and the endorsed perspective

causally influence their decision-making, while their personality traits play only a limited moderating role in this relationship.

Conclusion

In conclusion, the present study aligns with and replicates most of Nasello et al.'s findings (2023a; 2021). Approximately the same patterns of group-oriented choices were observed in the transplant and company's dilemmas: the participant's perspective and the type of trolley problem had a strong predictive effect on group-oriented decisions. The study demonstrated that causing harm, significant inconvenience, or being influenced by features such as the foot-in-the-door technique strongly influenced decision-making in group-oriented scenarios. Moreover, personality traits, specifically empathy and borderline traits, did not have the expected significant impacts on individuals' decision-making. Therefore, the present findings, combined with those of Nasello et al. (2023), suggest that the DPM (Greene et al., 2001; 2004) is an imperfect model for predicting moral decision-making. In fact, in the present study: (1) neither of the two main effects (the participant's perspective and the type of trolley problem) were significantly influenced by borderline personality disorder traits; and (2) affective empathy traits tend to positively predict group-oriented choices, which goes against the predictions made from the DPM.

In addition, we discovered that females diagnosed with borderline personality disorder exhibited distinct decision-making styles, characterized by a greater inclination towards avoidance and less inclination to logical seeking and assessment of alternative choices (rational style). We believe these may contribute to the development of specific borderline symptoms, suggesting that addressing these styles in therapy could have a meaningful impact. However, although the decision-making styles did not exert a significant predictive effect on moral decision-making, further investigation is needed to clarify the significant correlations found in this study. Concerning empathy, the findings are consistent with previous findings and a recent

meta-analysis showing that affective empathy has small associations with group-oriented decisions (Nasello & Triffaux, 2023), supporting the notion that empathy is not the most relevant predictor of moral decision-making (Decety, 2021; Decety & Cowell, 2014, 2015). In other words, there is probably a collective misconception around the idea that empathy explains all moral behaviors.

This study opens up new avenues for understanding the causal relationships among the different factors studied and highlights the potential moderating role of personality traits in these relationships. Furthermore, a certain consensus seems to be emerging regarding the need to review the theoretical foundations of the DPM.

Limitations

There are several limitations in the present study. The first one is that the study solely focuses on a female sample. Furthermore, we also established restrictive inclusion criteria regarding substance use history (alcohol and others), considering their potential impact on group-oriented choices (see Duke & Bègue, 2015). However, substance abuse is highly prevalent among individuals with borderline personality disorder. Finally, we did not analyze moral decision-making across genders, as females were exclusively presented with scenarios designed for females (in P_A and P_C). Therefore, it is not possible to claim that the results obtained from these perspectives are generalizable regardless of the target's gender.

Competing interests

The authors declare no competing interests.

Funding information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Approval

All procedures conducted in this study were carried out in compliance with the ethical standards set by the institutional research committee and adhered to the principles outlined in the 1964 Helsinki Declaration and its subsequent amendments, as well as the American Psychological Association's Ethical Principles in the Conduct of Research with Human Participants (2010). The "Hospitalo-facultaire" ethical committee of the CHU of Liège (Belgium) approved the study, reference n°: 2021-309.

Informed consent was obtained online from all participants included in the study.

Online data

The raw data file can be accessed via the following DOI link: https://osf.io/6tm98/

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington: American Psychiatric Association.

Bocéréan, C., & Dupret, E. A validation study of the Hospital Anxiety and Depression Scale (HADS) in a large sample of French employees. *BMC Psychiatry*, *14*, 354 (2014). https://doi.org/10.1186/s12888-014-0354-0

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413–425. https://doi.org/10.2307/1128984

Burghart, M., & Mier, D. (2022). No feelings for me, no feelings for you: A meta-analysis on alexithymia and empathy in psychopathy. *Personality and Individual Differences*, *194*, 111658. https://doi.org/10.1016/j.paid.2022.111658

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychological Bulletin, 135*(3), 495–510. https://doi.org/10.1037/a0015616

Curşeu, P. L., & Schruijer, S. G. (2012). Decision styles and rationality: An analysis of the predictive validity of the General Decision-Making Style Inventory. *Educational and Psychological Measurement, 72*(6), 1053-1062. https://doi.org/10.1177/0013164412448066

D'Ambrosio, F., Olivier, M., Didon, D., & Besche, C. (2009). The basic empathy scale: A French validation of a measure of empathy in youth. *Personality and Individual Differences, 46*(2), 160–165. https://doi.org/10.1016/j.paid.2008.09.020

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://psycn et.apa.org/doi/10.1037/0022-3514.44.1.113

Decety, J. (2021). Why empathy is not a reliable source of information in moral decision making. *Current Directions in Psychological Science*, *30*(5), 425–430. https://doi.org/10.1177/2F096 37214 21103 1943

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science*, *9*(5), 525–537. https://doi.org/10.1177/17456 91614 545130

Decety, J., & Cowell, J. M. (2015). Empathy, justice, and moral behavior. *AJOB Neuroscience*, *6*(3), 3–14. https://doi.org/10.1080/21507 740.2015.1047055

DeShong, H. L., Mullins-Sweatt, S. N., Miller, J. D., Widiger, T. A., & Lynam, D. R. (2016). Development of a short form of the five-factor borderline inventory. *Assessment, 23*(3), 342–352. https://doi.org/10.1177/1073191115581475

Dinić, B. M., Milosavljević, M., & Mandarić, J. (2021). Effects of Dark Tetrad traits on utilitarian moral judgement: The role of personal involvement and familiarity with the

victim. *Asian Journal of Social Psychology*, *24*(1), 48–58. https://doi.org/10.1111/ajsp.12422

Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition, 134*, 121-127. https://doi.org/10.1016/j.cognition.2014.09.006

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

First M, Gibbon M, Spitzer R, Williams JBW, Smith Benjamin L. (1994). *Structured Clinical Interview for DSM-IV Personality Disorders (SCID-II)*. Washington: American Psychiatric Association.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. In P. Foot (Ed.), (1977) *Virtues and vices and other essays in moral philosophy* (pp. 19– 32). Basil Blackwell.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*(2), 195.

Gallucci, M. (2019). GAMLj: General analyses for linear models. [Jamovi module]. Scribbr. https://gamlj.github.io/.

Girard, A. J., Reeve, C. L., & Bonaccio, S. (2016). Assessing decision-making style in French-speaking populations: Translation and validation of the general decision-making style questionnaire. *European Review of Applied Psychology*, *66*(6), 325-333. https://doi.org/10.1016/j.erap.2016.08.001

Gleichgerrcht, E., Torralva, T., Rattazzi, A., Marenco, V., Roca, M., & Manes, F. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high

functioning autism. *Social Cognitive and Affective Neuroscience*, *8*(7), 780–788. https://doi.org/10.1093/scan/nss067

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/scien ce.1062872

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*(12), 517–523. https://doi.org/10.1016/S1364 -6613(02)02011 -9

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814– 834. https://doi.org/10.1037/0033-295X.108.4.814

Hare, R. D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior*, *23*(1), 25–54. https://doi.org/10.1177/0093854896023001004

Hare, R. D., & Neumann, C. S. (2009). Psychopathy: Assessment and forensic implications. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, *54*(12), 791–802. https://doi.org/10.1177/070674370905401202

Herpertz, S. C., Werth, U., Lukas, G., Qunaibi, M., Schuerkens, A., Kunert, H. J., Freese, R., Flesch, M., Mueller-Isberner, R., Osterheider, M., & Sass, H. (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Archives of General Psychiatry*, *58*(8), 737–745. https://doi.org/10.1001/archpsyc.58.8.737

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*, 307–316. https://psycn et.apa.org/doi/10.1037/h0027580

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas? *PLoS One*, *11*(7), e0154780. https://doi.org/10.1371/journ al.pone.0154780

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence*, *29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010

Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten*. Johann Friedrich Hartknoch.

McCloskey, K. D., Cox, D. W., Ogrodniczuk, J. S., Laverdière, O., Joyce, A. S., & Kealy, D. (2021). Interpersonal problems and social dysfunction: Examining patients with avoidant and borderline personality disorder symptoms. *Journal of Clinical Psychology*, *77*(1), 329-339. https://doi.org/10.1002/jclp.23033

Mill, J. S. (1861). Utilitarianism. In J. M. Robson (Ed.), *The collected works of John Stuart Mill. Essays on ethics, religion and society* (Vol. X, pp. 203–259). University of Toronto Press.

Nasello, J. A., & Triffaux, J.-M. (2023). The role of empathy in trolley problems and variants: A systematic review and meta-analysis. British Journal of Social Psychology, 00, 1– 29. https://doi.org/10.1111/bjso.12654

Nasello, J. A., Blavier, A., & Triffaux, J. M. (2023b). French adaptation of the Five-Factor Borderline Inventory-Short Form. Current Psychology, 42, 5886–5897. https://doi.org/10.1007/s12144-021-01878-2

Nasello, J. A., Dardenne, B., Blavier, A., & Triffaux, J. M. (2021). Does empathy predict decision-making in everyday trolley-like problems? Current Psychology, 1-14, 2966–2979. https://doi.org/10.1007/s1214 4-021-01566-1

Nasello, J. A., Dardenne, B., Hansenne, M., Blavier, A., & Triffaux, J. M. (2023a). Moral decision-making in trolley problems and variants how do participants' perspectives,

borderline personality traits, and empathy predict choices? The Journal of Psychology: Interdisciplinary and Applied, 1-21. https://doi.org/10.1080/00223980.2023.2206604 Bentham, J. (1789). *Introduction to the principles of morals and legislation*. Payne.

Oudman, E., van Stigt Thans, S., Montoya, E. R., & Postma, A. (2021). Moral reasoning, moral decision-making, and empathy in Korsakoff's syndrome. *Journal of Neuropsychology*, *15*(3), 462–476. https://doi.org/10.1111/jnp.12233

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*, 501. https://doi.org/10.3389/fpsyg.2014.00501

Patil, I., Larsen, E. M., Kichic, R., & Gleichgerrcht, E. (2021). Moral cognition about harm in anxiety disorders: The importance of experienced emotion. *Psychological Reports*, *124*(6), 2501–2523. https://doi.org/10.1177/00332 94120 964134

Patil, I., Melsbach, J., Hennig-Fast, K., & Silani, G. (2016). Divergent roles of autistic and alexithymic traits in utilitarian moral judgments in adults with autism. *Scientific Reports*, *6*(1), 1–15. https://doi.org/10.1038/srep2 3637

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373-1379.

Pfohl, B., Blum, N., St. John, D., McCormick, B., Allen, J., & Black, D. W. (2009). Reliability and validity of the Borderline Evaluation of Severity Over Time (BEST): A self-rated scale to measure severity and change in persons with borderline personality disorder. *Journal of Personality Disorders*, *23*(3), 281-293. https://doi.org/10.1521/pedi.2009.23.3.281

Salgado, R. M., Pedrosa, R., & Bastos-Leite, A. J. (2020). Dysfunction of empathy and related processes in borderline personality disorder: A systematic review. *Harvard Review of Psychiatry*, *28*(4), 238–254. https://doi.org/10.1097/HRP.0000000000000260

Sauer-Zavala, S., & Barlow, D. H. (2014). The case for borderline personality disorder as an emotional disorder: Implications for treatment. *Clinical Psychology: Science and Practice, 21*(2), 118. https://doi.org/10.1111/cpsp.12063

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818-831. https://doi.org/10.1177/00131644950550050

Sidgwick, H. (1874). *The methods of ethics*. Macmillan.

Smillie, L. D., Katic, M., & Laham, S. M. (2021). Personality and moral judgment: Curious consequentialists and polite deontologists. *Journal of Personality, 89*(3), 549–564. https://doi.org/10.1111/jopy.12598

Takamatsu, R. (2018). Turning off the empathy switch: Lower empathic concern for the victim leads to utilitarian choices of action. *PLoS One*, *13*(9), e0203826. https://doi.org/10.1371/journ al.pone.0203826

Takamatsu, R. (2019). Personality correlates and utilitarian judgments in the everyday context: Psychopathic traits and differential effects of empathy, social dominance orientation, and dehumanization beliefs. *Personality and Individual Differences*, *146*, 1–8. https://doi.org/10.1016/j.paid.2019.03.029

Takamatsu, R., & Takai, J. (2019). With or without empathy: Primary psychopathy and difficulty in identifying feelings predict utilitarian judgment in sacrificial dilemmas. *Ethics & Behavior*, *29*(1), 71–85. https://doi.org/10.1080/10508422.2017.1367684

The Jamovi Project. (2019). Jamovi, version 1.6.23 [Computer Software]. Scribbr. https://www.jamovi.org

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204–217. https://www.jstor.org/stable/27902416 https://doi.org/10.5840/monist197659224

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361-370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

Synthèse de l’étude 3 :

Cette dernière étude offre une réplication des patterns de choix aux différents dilemmes selon les différentes perspectives. En effet, entre les études 2 et 3, les pourcentages de réponses sont approximativement similaires au niveau des deux dilemmes repris dans ces études (les dilemmes de l’entreprise et du chirurgien). De plus, on retrouve le même effet prédicteur de l'empathie affective sur les choix utilitaires dans la perspective B du dilemme de l'entreprise : dans la situation où le participant incarne le représentant, il y a une tendance (effet marginal) à sauver l'emploi des cinq autres employés quand l'individu a des traits d'empathie affective plus élevés.

Dès lors, plusieurs résultats importants sont apparus : un effet causal majeur de la perspective du participant dans la prise de décision morale, un effet causal également du type de scénario et, enfin, un effet prédicteur (ici, marginal) de l'empathie affective sur les choix orientés vers le groupe.

Contrairement à ce qui était attendu, les personnes qui souffrent d'un trouble de la personnalité borderline ne présentent pas une propension moins élevée à réaliser des choix utilitaires, comparativement aux volontaires sains (faibles au niveau des traits borderlines). Par rapport à ce qui avait été obtenu dans l'étude 2, les traits borderlines ne prédisent pas significativement les choix utilitaires dans aucun dilemme de type trolley. De ce fait, les prédictions du DPM sont, à nouveau, en décalage avec les observations empiriques. Effectivement, d'une part, il s'agit d'une population considérée comme étant hyper-réactive sur le plan émotionnel. D'autre part, cette population a une certaine propension à présenter des scores plus élevés en empathie affective (dans l'étude 3, nous avons observé des corrélations marginales entre les traits borderlines et l'empathie affective ; $r \approx .25$). Toutefois, non seulement l'effet prédicteur des traits borderline sur les choix n'est pas significatif mais la tendance inverse semble se dessiner (principalement dans les dilemmes de l’entreprise et du départ anticipé) : les individus diagnostiqués comme borderline et les contrôles qui présentent des traits borderlines élevés semblent avoir une plus forte propension à réaliser des choix utilitaires, comparativement aux individus sains présentant des traits borderlines faibles.

En conclusion, l'apport de cette dernière étude soutient l'hypothèse que le DPM est probablement un modèle incomplet pour expliquer et prédire les choix moraux dans des dilemmes de type trolley.

Les nouveaux apports théoriques de l'étude 3

En outre, cette dernière étude a permis d'apporter plusieurs éclairages quant aux spécificités des personnes diagnostiquées borderline. Effectivement, on retrouve plusieurs caractéristiques qui les distinguent des individus contrôles (faibles et hauts au niveau des traits borderlines) : les patientes présentent des scores significativement plus élevés au niveau des comportements négatifs (e.g., actions comprenant le fait de se mutiler), des affects de colère dysfonctionnels, du profond découragement (dimension fortement associée aux affects dépressifs), des perturbations subjectives de l'identité, des perturbations comportementales (i.e., une impulsivité marquée) et affectives, de la fragilité (dimension fortement associée à la vulnérabilité), de l'oppositionalité (i.e., une faible compliance) et de la précipitation (i.e., un manque de délibération). On remarque également une nette vulnérabilité des personnes qui vont présenter des traits borderlines élevés (patientes et contrôles) à la dépression et à l'anxiété, comparativement à des contrôles présentant des scores faibles en traits borderlines.

Au niveau des styles de prise de décision, on constate que les patientes borderlines ont un style de prise de décision plus évitant (ce style implique des efforts effrénés pour éviter de prendre des décisions) mais surtout un style de prise de décision moins rationnel (ce style implique une recherche de logique et une évaluation des alternatives à un choix).

En effet, concernant le style de prise de décision "évitant", les patientes borderlines et les contrôles ayant des scores élevés en traits borderlines ont des scores similaires, mais ces individus diffèrent significativement des contrôles ayant des scores faibles en traits borderlines. En contraste, pour le style « rationnel », il y a une gradation des scores où chacun des groupes diffère significativement des autres : les individus contrôles ayant des scores faibles en traits borderlines présentent les scores les plus élevés, suivis des individus contrôles ayant des scores élevés en traits borderlines, enfin, les patientes borderlines présentent les scores les plus faibles.

Enfin, nous n'avons pas obtenu d'effet prédicteur significatif des styles de prise de décision sur les choix moraux dans les dilemmes de type trolley. Toutefois, nous avons

obtenu des corrélations significatives entre les styles de décision et les choix moraux au sein de perspectives spécifiques : (1) deux corrélations positives ont été obtenues entre le style spontané et les choix utilitaires au sein de la perspective A du dilemme de l'entreprise (i.e., quand la participante incarne la patronne) et entre le style évitant et les choix utilitaires au sein de la perspective A du dilemme du départ anticipé (i.e., quand la participante incarne l'amie) ; et (2) une corrélation négative a été observée entre le style dépendent et les choix utilitaires au sein de la perspective C (i.e., quand la participante est une des cinq patientes mourantes). Ces résultats relativement nouveaux restent à interpréter avec prudence étant donné la faible taille de l'échantillon (n = 43) et le fait que ces associations ne subsistent pas dans les analyses GMM.

En définitive, cette étude nous a permis de répliquer des résultats majeurs de nos propres travaux (Nasello et al., 2021a ; Nasello et al., 2023a), à savoir : (1) les effets majeurs du type de dilemme et (2) de la perspective du participant, (3) l'effet faible de l'empathie affective sur les choix utilitaires, et (4) de souligner les lacunes du DPM. Ces conclusions s'inscrivent autour d'un corpus de données d'auteurs qui soulignent l'incomplétude de ce modèle pour expliquer et prédire les choix et jugements moraux (Oudman et al., 2021 ; Duke & Bègue, 2015 ; Horne & Powell, 2020 ; Smillie et al., 2021 ; Decety, 2021). Dans ce cadre, nous suggérons que le contexte moral des scénarios et la perspective qu'incarnent les participants ont une influence causale sur les choix moraux où les traits de personnalité agissent comme un modérateur de cette relation plutôt qu'un déterminant à part entière.

Section 3 : Discussion générale

Préambule

Cette dernière section vise à mettre en perspective les trois études expérimentales au travers un dénominateur commun : le DPM comme outil de prédiction des jugements et choix moraux et comme modèle explicatif de ces derniers. Sur base de ce modèle de nombreux auteurs ont investigué des populations cliniques particulières ou des traits de personnalité spécifiques qui font intervenir les composantes centrales du modèle (e.g., l'empathie). Comme cela a été le cas dans ce présent travail (mesure des traits borderline et de l'empathie).

A travers nos trois études, nous avons pu constater un effet significatif prédicteur (plutôt faible) de l'empathie affective sur les choix moraux. Toutefois, aucune de nos études n'a mis en évidence un effet prédicteur significatif de l'empathie cognitive sur les choix moraux. D'ailleurs, rares sont les effets prédicteurs significatifs de l'empathie cognitive sur les choix moraux dans la littérature scientifique. De même pour des facettes d'autres facteurs pouvant entrer dans la catégorie « processus cognitif » (e.g., la dimension « pensée opératoire » de l'alexithymie), sur les choix moraux.

Ces effets sont interpellants car ils renvoient à deux positions diamétralement opposées sur le rôle que l'on donne à l'empathie dans la morale. En effet, des auteurs comme Frans de Waal ou Jonathan Haidt vont considérer que l'empathie ou les émotions sont l'un des principaux piliers de la morale. En contraste, Jean Decety va attribuer à l'empathie un rôle probablement nécessaire mais il va définir que ce facteur n'apparaît pas comme le plus pertinent pour prédire les jugements et choix moraux.

Dès lors, *in fine*, quelle est la nature et la magnitude des relations entre l'empathie affective et cognitive et les différents aspects de la morale que sont les jugements moraux, les prises de décision morales et ce qu'on nomme « l'inclinaison morale[7] » (ou encore, la préférence morale) ? C'est à travers cette question capitale (qui repose notamment sur le DPM) que nous avons conduit une revue systématique et méta-analyse de la littérature scientifique. L'objectif était d'intégrer un maximum d'études quantitatives publiées dans des revues à comité de relecture ayant investigué (en objectif primaire ou périphérique) l'association entre empathie et les trois champs de la morale précités.

[7] L'inclinaison morale est une manière d'appréhender l'utilitarisme/déontologisme en termes de disposition plutôt « trait » ou comme la préférence générale d'un individu vers un champ de la morale.

Les relations entre l'empathie et les différents champs de la morale

Sur un total de 661 articles scientifiques, 34 ont été retenus. Nous avons pu réaliser un total de six méta-analyses et analyser systématiquement la littérature scientifique pour permettre de clarifier les associations entre empathie affective et cognitive et les jugements moraux, les prises de décision morales et la préférence morale dans les dilemmes de type trolley. Les 34 études incluent au total 26,015 participants issus de la population générale, mais également 383 individus issus de populations plus spécifiques (e.g., populations cliniques ou experts).

Deux méta-analyses ont été réalisées pour chaque champ de la morale : une pour évaluer l'association entre l'empathie affective et les jugements moraux, l'autre ciblait l'empathie cognitive et les jugements moraux ; une pour évaluer l'association entre l'empathie affective et les choix moraux, l'autre ciblait l'empathie cognitive et les choix moraux ; enfin, deux méta-analyses ayant ciblé l'association entre la préoccupation empathique (une dimension d'empathie affective) et les paramètres utilitaire et déontologique. Toutes les études incluent dans cette recherche de littérature ont appréhendé l'empathie via des méthodes variées, ce qui fournit un très large spectre d'investigation autour de cette notion. Concernant les différents champs de la morale, de nombreux types de dilemmes ont été utilisés, ce qui brasse également ces différents champs de manière éclectique.

Le lecteur est tout de même invité à prendre en considération, comme précisé dans la publication qui suit (Nasello & Triffaux, 2023b), que les champs de la morale investigués demeurent dans le spectre du dilemme du trolley qui investigue l'utilitarisme. Dès lors, nous n'appréhendons pas la morale dans son intégralité. Néanmoins, les trois champs visés brassent, à notre connaissance, l'entièreté du spectre du dilemme du trolley.

Article: The role of empathy in trolley problems and variants: A systematic review and meta-analysis

Received: 22 November 2022 | Accepted: 9 May 2023

DOI: 10.1111/bjso.12654

the british psychological society
promoting excellence in psychology

ARTICLE

The role of empathy in trolley problems and variants: A systematic review and meta-analysis

Julian A. Nasello[1,2] | **Jean-Marc Triffaux**[1,3]

[1]Psychiatric Day Hospital "La Clé", Liège, Belgium

[2]Department of Clinical Psychology, University of Liège, Liège, Belgium

[3]Department of Psychiatry, Medicine, University of Liège, CHU of Liège, Liège, Belgium

Correspondence
Julian A. Nasello, Psychiatric Day Hospital "La Clé"; Boulevard de la Constitution 153, Liège 4020, Belgium.
Email: julian.nasello@gmail.com

Abstract

The role of empathy in morality is a subject of ongoing scientific debate due to the lack of systematic reviews and meta-analyses on this topic. To address this gap, we conducted a PRISMA-based systematic quantitative review to investigate the role of empathy in moral judgements, decision-making, and inclinations using trolley problems and variants, which are popular types of moral dilemmas that explore utilitarianism and deontology. We searched for articles in four databases (PsycINFO, Pubmed, WorldWideScience, and Scopus) and performed citation searches. Out of 661 records, we selected 34 that studied the associations between empathy and moral judgements, moral decision-making, and/or moral inclinations. Six meta-analyses and systematic reviews of these records consistently showed small to moderate associations between affective empathy and these moral parameters, particularly in personal moral dilemmas involving intentional harm (although some approaches highlighted more complex associations between these parameters). Regarding other empathy domains, most studies found limited or insignificant links between cognitive empathy domains and moral judgements, decision-making, and inclinations. We discuss the nuances and implications of these results.

KEYWORDS

empathy, judgement/decision-making, moral inclination, morality, trolley problems, utilitarianism

INTRODUCTION

Moral dilemmas play a crucial role in human life as they require individuals to engage in complex moral judgements and decision-making. The recent COVID-19 pandemic and the underprepared health services serve as a poignant example of the importance of moral dilemmas (e.g., see Baheti & Nayak, 2022; Craig et al., 2022; Mitchell & Attipoe, 2020; Shao, 2020). Throughout history, philosophers and, more

recently, scientists have sought to explain the mechanisms that underlie moral judgements and moral decision-making. There are various types of moral dilemmas, with the prisoner's dilemma and the trolley problem being some of the most well-known. This paper will primarily focus on the trolley problem and its various iterations.

The trolley problem was popularized by Foot in 1967 and presented a scenario in which you are a runaway tram driver with the ability to switch tracks. On your current track, there are five workers who, if nothing is done, will be struck and killed by the tram. However, on the alternate track, there is only one worker. So the dilemma arises when you must choose between doing nothing and letting the five workers die or steering the tram to hit and kill the isolated worker. In 1976, Thomson proposed a variant of Foot's original scenario, known as the Footbridge problem. In this variation, you are positioned on a bridge and must choose between pushing one obese worker to stop the runaway tram and save the five (or doing nothing and letting the five workers perish). When individuals choose to save the five workers, we characterize this as a *utilitarian* choice.

Philosophers such as Bentham (1789), Mill (1861), and Sidgwick (1874) proposed the *Greatest Happiness Principle* for moral actions. According to this principle, an action is morally permissible if it produces the greatest happiness for the greatest number of individuals, and it should aim to reduce pain and increase pleasure. In other words, *utilitarianism* is impartial and prescribes focusing on the consequences of a moral problem and the option that will maximize utility (utilitarianism is also known as a *consequentialist* position).

On the other hand, deontological authors, such as Kant (1785), consider that pleasure and pain cannot be quantified and that what matters most is the action itself and the intention of the moral agent rather than its consequences. Like utilitarians, Kant believed in the existence of universal moral principles and absolute moral imperatives, from which obligations or duties are derived. However, unlike utilitarians, he regarded intentions, actions, and the sovereignty of individuals as central (Benlahcene et al., 2018). In this sense, individuals must consider others' dignity and rights, regardless of the consequences of their actions, and should focus on the moral value inherent in the act itself. Therefore, when individuals choose to save one worker in trolley problems, authors call this a *deontological choice*.

Greene and his colleagues made significant contributions to the study of moral judgements by employing Foot and Thomson's-like trolley problems in their research. They developed the Dual-Process Model (DPM) to explain how individuals make moral judgements (Greene, 2007; Greene et al., 2001, 2004; Greene & Haidt, 2002). The authors differentiated Foot's trolley problems from Thomson's based on the emotional and cognitive responses they elicited. Foot's scenarios were found to be less influenced by emotional responses and more by cognitive processes (specifically control-related processes in the dorsolateral prefrontal cortex) than Thomson's scenarios, which elicited stronger socio-emotional responses and fewer cognitive processes. Therefore, Foot's trolley problems were described as *impersonal* scenarios, while Thomson's were categorized as up close and *personal* scenarios. Consequently, Greene et al. (2004) suggested that emotional responses lead individuals to disapprove of personal moral violations. However, if the benefits of a personal moral violation outweigh the cost, an increase in cognitive activity (particularly in the dorsolateral prefrontal cortex) can override the emotional responses, resulting in a preference for personal moral violations. In other words, the DPM proposed that utilitarian judgements occur more frequently when cognitive processes override socio-emotional responses, typically in impersonal scenarios. On the other hand, more deontological judgements occur more frequently when the saliency of socio-emotional responses is high and cognitive processes do not outweigh these emotional responses, typically in personal scenarios.

It is worth briefly considering the terminology used in the fields of utilitarianism and deontology, as this has recently come under criticism from several authors (Nasello et al., 2021, 2023; Kahane, 2015; Kahane et al., 2015). For example, Kahane et al. (2015) found that the tendency to support harmful actions in order to save a larger group was not associated with, and may even be inversely related to, key markers of utilitarian concern for the greater good. This lack of correlation persisted even when the justification for these views was based unequivocally on utilitarian principles. Additionally,

'utilitarian' responses were found to be linked with characteristics, attitudes, and moral judgements that run counter to the impartial consideration for the collective benefit that is central to utilitarian ethics. These responses were associated with traits such as primary psychopathy, attitudes of rational egoism, and a permissive stance on clear moral violations (Kahane et al., 2015). As a result, some scholars have argued for more descriptive terminology, such as group-oriented choices instead of utilitarian choices, to describe situations where individuals opt to save five people (and individual-oriented choices, or even self-oriented choices for non-utilitarian choices, see Nasello et al., 2021, 2023). However, given that this terminology is relatively recent, and this positioning has not yet been widely adopted, we will use the traditional labels (i.e., deontological/utilitarian judgements, choices, or inclinations) in this article for better clarity.

Utilitarianism and deontology

Traditionally, studies utilize various trolley problems and their variations to evaluate utilitarian/deontological moral decision-making or moral judgement. In moral decision-making, participants must choose one option from a set of mutually exclusive options, whereas they assess the appropriateness of one or multiple options in moral judgement. However, there is often confusion between moral judgement and moral decision-making when using moral dilemmas. As Nasello et al. (2023) mentioned, making a decision marks the end of a sequential deliberation process, where one option is chosen among mutually exclusive options. In contrast, judgement is an evaluative process that involves assigning positive or negative evaluations to actions or characteristics of an individual in a given situation (Haidt, 2001). Thus, while moral judgement is a part of the sequential deliberation process and likely plays a significant role in moral decision-making, the degree of association between the two varies significantly based on several elements, such as dispositional traits (e.g., psychopathic traits) or inherent elements of the moral dilemma (Nasello et al., 2023). In fact, Nasello et al. (2023) noted that discriminating actions as good or bad can influence specific decisions, but not necessarily in a linear fashion. In other words, judging an action as bad does not necessarily prevent people from taking that action, demonstrating that judging the appropriateness of one action and deciding to take that action are two distinct operations.

Various approaches have been developed to investigate moral inclination, specifically utilitarian or deontological inclination viewed as a trait. Several research teams have created different methods for this purpose. In this paper, we will introduce two methods displayed in the retained records of this review (but see also Gawronski et al. [2017] for additional options). First, Kahane et al. (2015) developed "the Oxford Utilitarianism Scale," a 9-item scale that provides a total utilitarian score, which can be further divided into two domains: *permissive attitude toward instrumental harm* (i.e., the willingness to cause harm to achieve greater good) and *impartial concern for the greater good* (even at the cost of personal self-sacrifice). Second, Conway and Gawronski (2013) developed a methodology for obtaining scores of deontological and utilitarian inclinations using a set of ten congruent and ten incongruent moral dilemmas. The congruent dilemmas have a structure that leads to a consistent outcome between utilitarian and deontological positions (e.g., participants are asked whether it is acceptable to torture a man to discover and disarm harmless paint bombs; both utilitarian and deontological inclinations lead to rejecting torture). In contrast, the incongruent dilemmas are designed to lead to a different outcome between utilitarian and deontological positions (e.g., participants are asked whether it is acceptable to torture a man to discover and disarm deadly explosives placed around the city to save citizens' lives; deontologists should find the proposition unacceptable, while utilitarians should find it acceptable). Then, following two equations, we can calculate a *Deontological Parameter* ($DP = p(\text{unacceptable} \mid \text{incongruent})/(1 - UP)$) and a *Utilitarian Parameter* ($UP = p(\text{unacceptable} \mid \text{congruent}) - p(\text{unacceptable} \mid \text{incongruent})$; for a complete description, see Conway & Gawronski, 2013).

Emotional and cognitive processes in trolley problems and variants

Since its creation, the DPM (Greene, 2007; Greene et al., 2001) has received empirical support and has become a reference model in moral judgement studies. However, more recently, an increasing number of studies conducted on a wide range of emotional parameters (such as empathy, emotions, or alexithymia) and in various clinical populations (including psychopathic traits, borderline traits, and autistic traits) have questioned the importance given by the DPM to cognitive and emotional processes (Decety, 2021; Horne & Powell, 2016; Nasello et al., 2021; Smillie et al., 2020). In fact, the scientific community appears to be divided on this issue, with some influential authors arguing that empathy is likely essential but not the most relevant predictor of moral judgement and moral decision-making (Decety, 2021; Decety & Cowell, 2014, 2015), while others describe empathy as one of the two pillars[1] of human morality (de Waal et al., 2006). Similarly, there are the same dichotomic positions about emotions: Haidt (2003) suggested that emotions are in charge of the temple of morality, whereas other authors argued that there is insufficient evidence to support the idea that emotions play a significant role in moral cognition and moral judgements (Horne & Powell, 2016; May, 2018).

Empathy: definitions and measurements

Empathy is traditionally divided into two components: *affective empathy*, which pertains to the ability to experience the emotions of another (Bryant, 1982), and *cognitive empathy*, which refers to the ability to understand the emotions of others (Hogan, 1969), although this capacity to comprehend is sometimes extended to all mental states (Davis, 1980, 1983). However, empathy is much more complex than this binary conception suggests. Cuff et al. (2016) identified over forty definitions of empathy in their systematic review, demonstrating the large range of perspectives on empathy. The authors proposed a definition to synthesize these various conceptions: "empathy is an emotional response (affective), dependent upon the interaction between trait capacities and state influences. Empathic processes are automatically elicited but are also shaped by top-down control processes. The resulting emotion is similar to one's perception (directly experienced or imagined) and understanding (cognitive empathy) of the stimulus emotion, with a recognition that the source of the emotion is not one's own." (Cuff et al., 2016, p. 150). More recently, Eklund and Meranius (2021) have described four main themes in the empathy literature: *understanding* (i.e., knowing something about the mental life of another person), *feeling* (i.e., an affective response to another person's situation), *sharing* (i.e., experiencing similar states to those experienced by the other person), and *self-other differentiation* (i.e., recognizing a differentiation between oneself and the other person).

Over the years, several empathy scales have been developed to capture this complex phenomenon. The most commonly used in research is probably the Interpersonal Reactivity Index (IRI; Davis, 1983), but other scales include the Empathy Quotient (Baron-Cohen & Wheelwright, 2004), the Basic Empathy Scale (BES; Jolliffe & Farrington, 2006), and the Hogan Empathy Scale (Hogan, 1969). In this section, we will describe the IRI and the BES because most of the research selected in this systematic review and meta-analysis used these questionnaires.

The IRI is a 28-item questionnaire that assesses four components: (1) *perspective-taking*, which refers to the ability to adopt another's perspective or point of view; (2) *fantasy*, which refers to people's propensity to get involved in fictional situations and identify with fictional characters in books, movies, or plays. These are two components of cognitive empathy. The other two domains are (3) *empathic concern*, which refers to the respondent's tendency to experience feelings of concern or compassion for others, and (4) *personal distress*, which refers to the tendency to experience distress

[1]The other pillar is reciprocity.

or discomfort in response to others' emotional distress. These are two domains of affective empathy. The IRI has several advantages: it is a multidimensional scale with good psychometric properties that is easy to use and quick to complete.

On the other hand, the BES was created to address the limitations of the IRI. According to Jolliffe and Farrington (2006), empathy is limited to emotional abilities, not other mental states such as thoughts or intentions. Therefore, the BES focuses on four basic emotions (anger, fear, happiness, and sadness) in its two empathy components. This 20-item scale measures *affective empathy* (i.e., the capacity to experience the emotions of another) and *cognitive empathy* (i.e., the capacity to comprehend the emotions of another). Like the IRI, the BES displays good psychometric properties, is easy to use, and is quick to complete.

Objective

Regrettably, there is currently a lack of systematic reviews and meta-analyses that can provide clarity on the role of both emotional and cognitive processes in moral judgements, moral decision-making, and moral inclination. Although this is an extensive area of research that requires encouragement in the scientific community, we developed here a systematic review and meta-analysis that focuses on the links between empathy[2] (a phenomenon that encompasses both emotional and cognitive aspects) and moral judgements, moral decision-making, and moral inclination in trolley problems and their variants.

METHOD

Literature search and study selection

The study selection process was carried out in accordance with the PRISMA guidelines (Moher et al., 2009; Page et al., 2021) to investigate the associations between empathy and trolley problems and their variants. A comprehensive search of peer-reviewed studies published before July 30, 2022, was conducted using four databases: PubMed, PsycINFO, WorldWideScience, and Scopus. The search strategy included terms related to empathy and trolley problems, which could be found in titles, abstracts, keywords, subject headings, or the full text. The following terms were searched in the four databases: (trolley) OR (sacrificial dilemma*) AND (empath*). In addition, citation searches were used to supplement the references selected in this systematic review and meta-analysis (see Figure 1).

Inclusion and exclusion criteria

The present review included references that met the following criteria: (1) the reference must be an empirical study (theoretical papers, systematic reviews, or meta-analyses were excluded); (2) the reference must be published in a peer-reviewed journal; (3) the title, abstract, or keywords of the reference should include the terms selected in this systematic review and meta-analysis; (4) the study must have used an adult sample (18+ years); and (5) the study should be published in the English language. Studies that were excluded had the following criteria: (a) they were qualitative studies; (b) they were theoretical

[2]Important note: The current paper does not investigate the associations between psychopathic traits and moral judgements, decision-making, or moral inclination. Although psychopathy is characterized by a lack of affective empathy and a deficiency in perspective-taking (see Burghart & Mier, 2022 for a meta-analysis), this psychopathology is far more complex than just empathy impairments. Other factors, such as marked impulsivity and antisocial traits, also characterize psychopathy. Since empathy impairments are only a small part of this complex pathology, we did not include psychopathic traits in the present analyses.

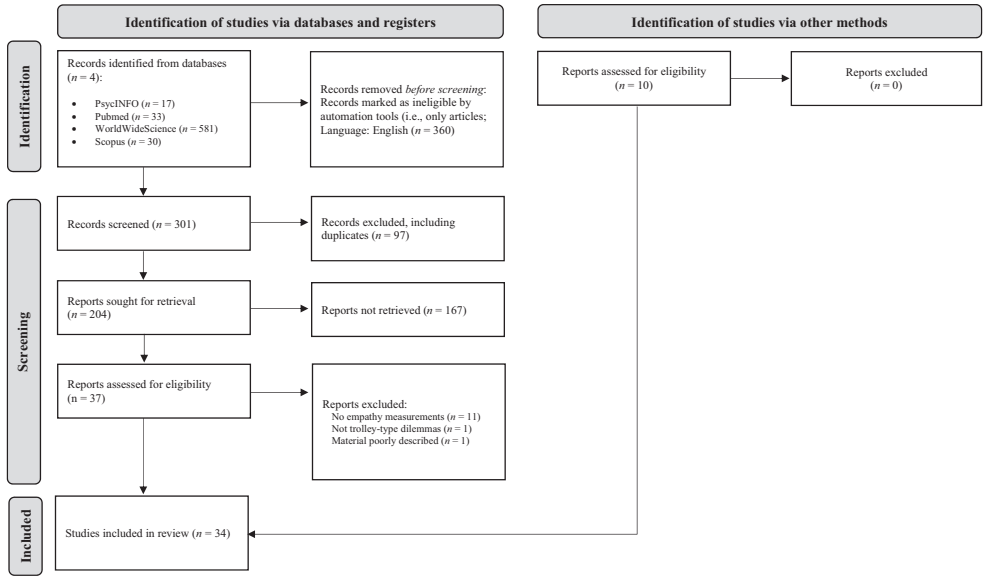


FIGURE 1 PRISMA flow diagram.

papers; or (c) they were reviews. No restrictions were placed on the publication date or the empathy measurement, as our systematic review and meta-analysis aimed to obtain a comprehensive understanding of the role played by empathy in moral judgements, moral decision-making, and moral inclination in trolley problems and variants.

Data extraction

A total of 661 records were retrieved: 17 from PsycINFO, 33 from Pubmed, 581 from WorldWideScience, and 30 from Scopus, with limits applied for humans, the English language, and age 18+ years. After removing duplicates and carefully examining titles, abstracts, and full texts, 24 references were included. An additional ten articles were identified through citation searches. Therefore, a total of 34 papers were included in this systematic review and meta-analysis (see Figure 1), with all studies examining the associations between empathy and moral judgements, moral decision-making, or moral inclinations using trolley problems and related variants, either as a primary or secondary objective. Note that this study was not preregistered.

Analysis

Table 1 displays the following information: (1) the authors' names and date of publication, (2) the sample size and some descriptive information, (3) the utilitarianism/deontology measurement, (4) whether the authors had studied moral judgements, moral decision-making, moral inclination or specific combinations, (5) the empathy measurement used in the study, (6) the statistics performed in the study, (7) the main results of the study regarding the scope of this review, and (8) the effect sizes of the significant results. Importantly, it is worth noting that we converted all retained records into numbers (1–34; see Table 1) to avoid complete citations in the text.

We conducted a meta-analysis when we had a $k \geq 9$ number of studies to determine the overall estimated association between empathy domains and moral judgements, moral decision-making, and moral inclination. When the number of studies was $k < 9$, we systematically reviewed the records studying those associations.

TABLE 1 Compiled information from retained records.

#	Authors and date	Population	Trolley types or measurements	MJ, MI, or MDM	Empathy measurements	Statistics	Main results	Effect sizes
1.	Baez et al. (2017)	**Study 2**: N = 334 (164 women); M_{age} = 46.6; SD = 10.6 Gen. pop. (students and teachers; mostly south American participants).	Footbridge (personal dilemma) & Foot's Trolley problem (impersonal dilemma)	MDM	IRI (EC, PD, PT, FS)	Multivariate regressions (DV = Empathy domains; IVs = MDM in personal dilemma & Gender); MDM: UC = 1	UC → EC: (*beta* = -0.16; p < .001) UC → PD: (*beta* = -0.01; p = .92) UC → FS: (*beta* = -0.08; p = .09) UC → PT: (*beta* = -0.01; p = .80)	η^2 = 0.02
2.	Behnke et al. (2020)	N = 1004 (534 women); M_{age} = 24.5; IQR = 4.6 Gen. pop. (mostly german high-school graduate students).	5 personal moral dilemmas: - Lifeboat - Footbridge - Transplant - Crying baby - Sophie's choice	MJ (appropriateness)	IRI (EC, PD, PT, FS) Some personality traits and IRI domains were combined to form five new factors (see footnote [6]). Regarding empathy domains, the main factors were PF1 (including EC, PT, & FS) and PF2 (including PD).	Correlations Linear mixed regressions models (LMRM) (DV = Perceived moral appropriateness; IVs = PF1-PF5, motive for killing, avoidability, and interactions between all these variables)	Footbridge: PF1 – Appropriateness (r = -.11; p < .001); PF2 – Appropriateness (r = -.01; *ns.*). Lifeboat: PF1 – Appropriateness (r = -.08; *ns.*); PF2 – Appropriateness (r = -.05; *ns.*). Transplant & Footbridge (LMRM): PF1 (B = -0.24; SE = 0.04; p < .001); PF2 (B = 0.00; SE = 0.04; p = .932). Crying baby & Sophie's choice (LMRM): PF1 (B = -0.05; SE = 0.06; p = .352); PF2 (B = 0.07; SE = 0.05; p = .177).	Small η^2p = .01
3.	Bostyn et al. (2018)	N = 192 Gen. pop. (students).	10 hypothetical trolley-style dilemmas (personal moral dilemmas) 1 real-life version of the trolley problem (using mice) 1 hypothetical version of the mouse dilemma	MJ (appropriateness) & MDM for a real-life trolley-type dilemma (using mice)	IRI (EC & PT) Animal Empathy Scale (Paul, 2000) Empathy for mice	Correlations (Consequentialism appropriateness: CA) (Deontology appropriateness: DA) Logistic regression (DV = Consequentialist decision; IVs = Need for cognition, Empathic concern, Perspective Taking, Primary Psychopathy, Moral identity: identification, Moral identity: symbolism, Animal empathy, Empathy for mice, Gender, and Age)	CA – EC (r = -.14; p < .05); CA – PT (r = .03; *ns.*); CA – Animal empathy (r = -.03; *ns.*); CA – Mice empathy (r = -.14; p < .05); Real – EC (r = .071; *ns.*) Real – PT (r = .019; *ns.*) DA – EC (r = -.12; p < .10); DA – PT (r = -.08; *ns.*); DA – Animal empathy (r = -.07; *ns.*); DA – Mice empathy (r = -.05; *ns.*); Logistic regression for the real-life dilemma: Empathy measures (*ns.*)	Small Small
4.	Côté et al. (2013)	**Study 2**: N = 229 (130 women); M_{age} = 33.8; SD = 11.9 Gen. pop. (Mturk workers, US residents, mostly Caucasians). **Study 3**: N = 91 (55 women); M_{age} = 34.7; SD = 12.3 Gen. pop. (Mturk workers, US residents, mostly Caucasians).	Utilitarian choices in an allocation money task (taking dollars from a target to benefit three other people)	MDM	Study 2: Adapted measure from Kogut & Ritov (2005) - Compassionate - Sympathy - Worried - Upset - Sad Study 3: three-item scale (Oveis et al., 2010): - Compassionate - Moved - Sympathy & Perspective-taking induction for the losing member (*"As you make your decision, think about the feelings and the well-being of the 'lose member' of the group. Concentrate on trying to imagine how the 'lose member' feels and how your decision will influence him or her. Try to feel the impact of your decision on how the 'lose member' of your group will feel."*)	Regressions (DV = Utilitarian Choice [UC]; IVs = Gender, Age, Ethnicity, Religiosity, Conservative political orientation, Social class, Empathy) Correlations	**Study 2:** Empathy → UC: B = -0.36; SE = 0.06; p < .001 Empathy – UC: (r = -.41; p < .001) **Study 3:** Perspective-Taking induction → UC: B = 0.11; SE = 0.03; *ns.* But inducing PT extinguished the association between social class and UC: B = -0.08; SE = 0.12; *ns.*	R^2 = .11 Moderate
5.	Conway et al. (2018)	**Study 2**: N = 172 (67 women); M_{age} = 34.4; SD = 9.92 Gen. pop. (MTurk workers). **Study 3**: N = 182 (86 women); M_{age} = 33.4; SD = 10.48 Gen. pop. (MTurk workers). **Study 5**: N = 171 (80 women); M_{age} = 34.3; SD = 10.9 Gen. pop. (MTurk workers).	10 trolley problem variants: incongruent vs. congruent dilemmas (Conway and Gawronski, 2013)	MJ (appropriateness of the harmful action [Utilitarian action]) & Moral Inclination	IRI (EC)	Correlations MJ: higher scores = more acceptance of the utilitarian action; lower scores = more rejection of the utilitarian action (=deontological judgment).	**Study 2:** UJ – EC (r = -.20; p < .01) DP – EC (r = .29; p < .01) UP – EC (r = -.05; *ns.*) **Study 3:** UJ – EC (r = -.18; p < .05) DP – EC (r = .17; p < .05) UP – EC (r = -.06; *ns.*) **Study 5:** UJ – EC (r = -.25; p < .001) DP – EC (r = .24; p < .01) UP – EC (r = .02; *ns.*)	Small Small/Mod. Small Small Small Small
6.	Dinic et al. (2021)	N = 758 (477 women); M_{age} = 26.6; SD = 7.292 Gen. pop. (mostly university students)	4 moral dilemmas (Bartels and Pizarro, 2011): personal and impersonal moral dilemmas	MDM (*Would you do..?*)	The Affective and Cognitive Measure of Empathy (Vachon & Lynam, 2016): Cognitive empathy (CE) Affective Resonance (AR) Affective Dissonance (AD) CE refers to the detection and understanding of other people's emotions and feelings [it might be compared to perspective-taking]. AR is defined as the emotional influence and feeling of sympathy, compassion, pity, and concern for someone else's misfortune [it might be compared to empathic concern]. AD refers to antagonizing emotional reactions and responses (e.g., feeling enjoyment or pleasure from others' misfortune and pain).	Correlations Hierarchical MLR (DV = Utilitarian Action; IVs = age, gender, empathy domains, HEXACO domains, Machiavellianism, Psychopathy, Sadism, Narcissism)	Impersonal dilemmas: UC – CE (r = -.03; *ns.*) UC – AR (r = -.18; p < .001) UC – AD (r = -.20; p < .001) Personal dilemmas: UC – CE (r = -.02; *ns.*) UC – AR (r = -.25; p < .001) UC – AD (r = -.32; p < .001) HRA: Impersonal dilemmas: CE → UC (β = -.01; *ns.*) AR → UC (β = .02; *ns.*) AD → UC (β = -.12; p < .05) Personal dilemmas: CE → UC (β = .03; *ns.*) AR → UC (β = -.01; *ns.*) AD → UC (β = -.16; p < .001)	Small Small Small Moderate ΔR^2 = .04* ΔR^2 = .09* *For all empathy domains
7.	Gleichgerrcht & Young (2013)	**Study 1**: N = 1339 (752 women); M_{age} = 25.7; SD = 11.2 Gen. pop. (around 13 education years). **Study 2**: N = 896; M_{age} = 24.8; SD = 15.7 Gen. pop. (around 13 education years). **Study 3**: N = 513; M_{age} = 25.7; SD = 12.9 Gen. pop. (around 13 education years).	**Study 1:** Footbridge (personal dilemma) & Foot's Trolley problem (impersonal dilemma) **Study 2:** Crying baby (personal dilemma) & Standard fumes dilemma (impersonal dilemma) **Study 3:** Transplant dilemma (personal dilemma)	MDM (*Would you do..?*)	IRI (EC, PD, PT, FS)	Student's *t*-tests (differences between U and non-U responders) Discriminant analyses (Variables: age, gender, education, Moral behavior inventory, Daily Spiritual Experience Scale, and IRI domains)	**Study 1:** Impersonal scenario: PT: $t_{(1332)}$ = 0.70; p = .49 FS: $t_{(1332)}$ = 1.08; p = .28 EC: $t_{(1332)}$ = -1.73; p = .08 PD: $t_{(1332)}$ = 0.06; p = .95 Personal scenario: PT: $t_{(1332)}$ = 0.93; p = .35 FS: $t_{(1332)}$ = -1.75; p = .08 EC: $t_{(1332)}$ = -9.27; p < .001 PD: $t_{(1332)}$ = -0.77; p = .44 **Study 2:** Impersonal scenario: PT: $t_{(883)}$ = 0.51; p = .61 FS: $t_{(883)}$ = 0.94; p = .35 EC: $t_{(883)}$ = -1.26; p = .21 PD: $t_{(883)}$ = -0.44; p = .66 Personal scenario: PT: $t_{(883)}$ = -1.33; p = .19 FS: $t_{(883)}$ = -1.82; p = .07 EC: $t_{(883)}$ = -6.79; p < .001 PD: $t_{(883)}$ = -0.74; p = .46 **Study 3:** Personal scenario: PT: $t_{(506)}$ = -0.15; p = .88 FS: $t_{(506)}$ = -1.22; p = .11 EC: $t_{(506)}$ = -7.18; p < .001 PD: $t_{(506)}$ = -0.58; p = .56 Discriminant function (**study 1 only**): EC was the factor that most strongly classified utilitarian participants apart (.84)	d = .55 d = .46 d = .64

(Continues)

TABLE 1 (Continued)

#	Authors and date	Population	Trolley types or measurements	MJ, MI, or MDM	Empathy measurements	Statistics	Main results	Effect sizes
8.	Gleichgerrcht et al. (2013)	Adults with high-functioning autism: N = 36 (12 women) M_{age} = 32.6; SD = 10.9 (around 15 education years). HC: N = 36 (13 women); M_{age} = 34.2; SD = 8.7 (around 15 education years).	Footbridge (personal dilemma) & Foot's Trolley problem (impersonal dilemma)	MDM (also MJ [appropriateness], but no investigations on the associations between MJ and empathy were conducted)	EQ IRI (EC, PD, PT)	Chi-square tests Student's t-tests (differences between UC and DC)	Impersonal scenario: EQ: DC = UC ($t_{(34)}$ = 1.94; p = .06) PT: $t_{(31)}$ = 0.07; p = .95 EC: $t_{(31)}$ = 0.47; p = .65 PD: $t_{(31)}$ = 1.61; p = .09 Personal scenario: EQ: $t_{(34)}$ = 1.18; p = .26 PT: $t_{(31)}$ = 2.09; p = .04; DC > UC EC: $t_{(31)}$ = 0.64; p = .53 PD: $t_{(31)}$ = 1.12; p = .28	Unreported.
9.	Hannikainen et al. (2018)	**Study 1, wave 1**: N = 4134 (1162 women); M_{age} = 30.3; SD = 13.1 Gen. pop. (college or graduate students, mostly US nationals, with no religious affiliation). **Study 2, wave 2**: N = 123 (50 women); Mdn_{age} = 41 Gen. pop. (participants from stud 1: mostly US nationals with no religious affiliation, who held a master's or bachelor's degree). **Study 3, waves 1 & 3**: N = 6754; *Age* (Wave 1: M_{age} = 28.6; SD = 11.8; Wave 3: M_{age} = 28.7; SD = 11.7). Gen. pop. (students, mostly US nationals, with no religious affiliation).	13 personal dilemmas	MJ (permissibility of the utilitarian action)	IRI (EC, PD, PT, FS)	Correlations Multiple linear regression (all IRI domains as IVs; DV = MJ) Hierarchical MLR (DV = MJ; IVs = Wave; age, gender, religious affiliation, educational attainment, all IRI domains, and religiosity)	**Study 1:** UJ – FS: r = 0.0 (*ns.*) UJ – PT: r = -0.11 (p < .005) UJ – EC: r = -0.29 (p < .005) UJ – PD: r = -0.08 (p < .005) MLR: EC: B = -0.58; t = -18.1; p < .001 **Study 2 (longitudinal study, after eight years):** No shift in MJ: $t_{(122)}$ = 1.11; p = .250 MLR: EC (study 1): B = -0.69; t = -4.00; p < .001; EC (study 2): B = -0.50; t = -3.04; p = .003. **Study 3:** Hierarchical MLR: EC: *Final* B = -0.36; p < .001 PD: *Final* B = -0.13; p < .001 FS: *Final* B = 0.10; p < .05 PT: *Final* B = -0.02; *ns.*	Small Small/Mod. Small η^2 = 0.087 $\eta_p{}^2$ = 0.15 $\eta_p{}^2$ = 0.08 $\Delta R^2_{adj.}$ = 0.07* *all empathy domains with religiosity.
10.	Kahane et al. (2018)	**Study 3**, Moral philosophy experts: N = 81 (23 women) M_{age} = 32; SD = 9.7 (mostly graduate students or post-doctoral researchers).	Oxford Utilitarianism Scale: - Permissive attitude toward instrumental harm (OUS-IH) - Impartial concern for the greater good (OUS-IB)	Moral inclination	IRI (EC)	Correlations	EC – OUS-IB: r = 0.33 (p < .005) EC – OUS-IH: r = -0.16 (p < .005)	Moderate Small
11.	Kahane et al. (2015)	**Study 1**: N = 194 (66 women); M_{age} = 31; SD = 9.49 Gen. pop. (Mturk workers).	4 personal dilemmas (from Moore et al., 2008): - Footbridge - Epidemic - Soldiers - Hospital	MJ (*Wrongness of utilitarian action*) & MDM	IRI (EC)	Correlations Multiple linear regression (DV = Wrongness of UC; IVs = primary psychopathy and EC)	**Study 1:** EC – UC: r = -0.14 (p = .06) EC – UJ: r = 0.17 (p < .05) MLR: EC: *ns.*	Small
12.	Maranges et al. (2022)	**Study 1**: N = 116 (90 women); M_{age} = 19.8; SD = 1.21 Gen. pop. (mostly white students). **Study 2**: N = 247 (131 women); M_{age} = 35.5; SD = 10.4 Gen. pop. (Mturk workers, mostly white)	10 moral dilemmas (incongruent vs. congruent) see Conway & Gawronski, 2013. - Deontology parameter (DP) - Utilitarian parameter (UP)	MJ (appropriateness of the harmful action [Utilitarian action]) & Moral Inclination	IRI (EC)	Correlations Multiple linear regression (MLR1: DV = DP; IVs = EC, discomfort with helping, need to belong, emotion control, concern for the individual, concern for the group; controlling variables = age and gender) (MLR2: DV = UP; IVs = EC, discomfort with helping, need to belong, emotion control, concern for the individual, concern for the group; controlling variables = age and gender) Path analysis (DVs = DP and UP; Regressors = Avoidant attachment, anxious attachment, empathic concern, need to belong, emotion control, concern for the individual, concern for the group)	**Study 1:** EC – UJ: r = -0.16 (p < .10) EC – DP: r = 0.24 (p < .05) EC – UP: r = 0.02 (*ns.*) **Study 2:** EC – UJ: r = -0.18 (p < .01) EC – DP: r = 0.20 (p < .05) EC – UP: r = 0.11 (*ns.*) MLR1: EC → DP: B = .02; SE = .12; p = .850 MLR2: EC → UP: B = -.09; SE = .12; p = .468 Path analysis: EC → DP (.21; p < .01); EC → UP (.20; p < .01).	Small Small Small Unreported. Unreported.
13.	Maranges et al. (2021)	**Study 4**: N = 250 (121 women); M_{age} = 34.6; SD = 10.2 Gen. pop. (Mturk workers, mostly white)	20 moral dilemmas (ten incongruent vs. ten congruent) using Conway & Gawronski's methodology (2013) - Deontology parameter (DP) - Utilitarian parameter (UP)	Moral inclination	IRI (EC)	Correlations	**Study 4:** EC – DP: r = 0.29 (p < .001) EC – UP: r = 0.24 (p < .01)	Small/Mod. Small
14.	Nasello et al. (2021)	N = 255 (189 women); M_{age} = 20.9; SD = 2.90 Gen. pop. (university students).	2 everyday trolley-like problems (1 impersonal and 1 personal)	MDM	BES (affective and cognitive empathy; AE, CE)	Correlations [UC = 0; DC = 1] General Mixed Model analysis (DV = UC; IVs = trolley problem scenarios, gender, age, affective empathy, cognitive empathy, scenarios * AE, gender * AE, scenarios * CE, gender * CE, scenarios * gender * AE, scenarios * gender * AE)	Impersonal scenario: AE – UC: r = 0.055 (*ns.*) CE – UC: r = 0.092 (*ns.*) Personal scenario: AE – UC: r = 0.038 (*ns.*) CE – UC: r = 0.077 (*ns.*) GMM: AE → UC: B = 0.021; SE = 0.024; p = .394; CE → UC: B = -0.05; SE = 0.036; p = .164 Scenario * Gender * AE → UC: B = 0.190; SE = 0.081; p = .019 Scenario * Gender * CE → UC: B = -0.044; SE = 0.124; p = .725	Exp(B) = 1.21
15.	Nakamura et al. (2014)	**Study 2**: N = 41 (15 women); M_{age} = 19.7; SD = 1.12 Gen. pop. (Japanese undergraduate students).	16 moral dilemmas: - 8 high-conflict (Crying baby, Footbridge, Sophie's choice, Vitamins, Lawrence of Arabia, Submarine, Bomb, and Preventing the spread) - 8 low-conflict (Country road, Plane crash, Hired rapist, Infanticide, Architech, Hard times, Transplant, Smother for dollars)	MJ (appropriateness)	Four empathy-related adjectives: sympathetic, compassionate, softhearted, and tender (Zhou et al., 2012)	Spearman's correlations Path analysis DV = Moral acceptability of high-conflict moral dilemmas; Regressors = Temperature cue, Inclusion of Other in the Self scores (Aron et al., 1992), empathy)	**Study 2:** High conflict Empathy – UJ: r = -0.424 (p < .01) Low conflict Empathy – UJ: r = -0.018 (*ns.*) The direct link between empathy and acceptability of the utilitarian option in high-conflict dilemmas (personal dilemmas): β = -0.33; p = .022	Moderate R^2 = 0.073
16.	Oudman et al. (2021)	Korsakoff's patients: N = 20 (8 women); M_{age} = 61.8; SD = 7.83; education level: M = 4.5 (1.24). HC: N = 20 (8 women); M_{age} = 64.1; SD = 9.44; education level: M = 4.95 (0.95).	10 moral dilemmas from Kahane et al.(2012)	Moral inclination: Utilitarian Moral Reasoning (UMR)	IRI (EC & PT)	Multiple linear regression MLR1: DV = UMR; IVs = level of education, age, gender, PT, EC, FAB, Group (KP or HC) MLR2: DV = DMR; IVs = level of education, age, gender, PT, EC, FAB, Group (KP or HC)	MLR1: PT → UMR: β = 0.00; p = .81 EC → UMR: β = 0.04; p < .01 MLR2: No significant predictors were reported.	Unreported.

TABLE 1 (Continued)

#	Authors and date	Population	Trolley types or measurements	MJ, MI, or MDM	Empathy measurements	Statistics	Main results	Effect sizes
17.	Patil & Silani (2014)	N = 331 (215 women); M_{age} = 24.1; SD = 5.50 Gen. pop. (Italian speakers).	2 trolley problems (one impersonal [Standard fumes] and one personal [Footbridge])	MJ (appropriateness)	IRI (EC, PD, PT, FS)	Multiple linear regression DV = UJ; IVs = IRI domains; Criterion Variables = Impersonal, Personal dilemmas. Correlations	MLR: Personal dilemmas: EC → UJ: B = -0.067; p = .009 Impersonal dilemma: EC – UJ: r = -0.076 (*ns.*) PD – UJ: r = -0.011 (*ns.*) PT – UJ: r = 0.003 (*ns.*) FS – UJ: r = -0.008 (*ns.*) Personal dilemma: EC – UJ: r = -0.119 (p < .05) PD – UJ: r = 0.0033 (*ns.*) PT – UJ: r = 0.043 (*ns.*) FS – UJ: r = 0.053 (*ns.*) None of the other empathy domains significantly predicted UJ in personal or impersonal dilemmas. No other significant correlation was found between empathy domains and UJ in personal or impersonal dilemmas.	Unreported. Small
18.	Patil et al. (2021)	Anxiety disorders patients: N = 95 (48 women); General M_{age} = 37.1 HC: N = 63 (26 women); M_{age} = 33.9; SD = 11.2	2 trolley problems: 1 impersonal (Standard fumes) and 1 personal (Footbridge)	MJ (appropriateness) (also MDM, but no investigations on the associations between MDM and empathy were performed)	IRI (EC, PT, PD)	Correlations (Holm's adjustment p values)	Impersonal dilemma: EC – UJ: r = -0.16 (*ns.*) PD – UJ: r = -0.03 (*ns.*) PT – UJ: r = -0.07 (*ns.*) Personal dilemma: EC – UJ: r = 0.04 (*ns.*) PD – UJ: r = 0.08 (*ns.*) PT – UJ: r = -0.14 (*ns.*)	
19.	Rosas et al. (2019)	**Study 1**: N = 230 (69 women); M_{age} = 21 Gen. pop. (students). **Study 2**: N = 590 (181 women); M_{age} = 22.1 Gen. pop. (students).	**Study 1**, 8 personal moral dilemmas: - Park - Footbridge - Selfish Flames - Selfish Explosives - Doomed Guide - Doomed Boat - Assistant 100k - Dam **Study 2**, 6 moral dilemmas: - Aquarium - Footbridge - Assistant 100k - Vaccine - Sacrifice Comfort - Let Mother Die	MJ (Rightness/ Wrongness of the utilitarian action)	IRI (EC)	Correlations **Study 1:** - Strict (Park + Footbridge) = personal moral dilemmas - 100k (Assistant 100k + Dam) - Selfish (Selfish flames + selfish explosives) - Doomed (Doomed guide + Doomed boat) **Study 2:** - Strict (Aquarium + Footbridge) - 100k (Assistant 100k + Vaccine) Partial correlations (controlling primary psychopathy) Cluster analysis	**Study 1:** EC – Strict: r = -0.254 (p < .001) EC – Selfish: r = -0.212 (p = .001) EC – Doomed: r = -0.355 (p < .001) EC – 100k: r = -0.140 (p = .034) **Study 2:** EC – Strict: r = -0.09 (p = .030) EC – Strict: r_p = -0.013 (*ns.*) EC – 100k: r = -0.138 (p = .001) EC – 100k: r_p = -0.046 (*ns.*) EC – Sacrifice comfort: r = 0.384 (p < .001) EC – Sacrifice comfort: r_p = 0.241 (p <.001) EC – Let mother die: r = 0.117 (p = .034) EC – Let mother die: r_p = 0.045 (*ns.*) Cluster analysis: All UJ responders displayed higher EC scores compared to all DJ responders (17.2, SD = 4.70; 16.3; SD = 4.26, respectively)	Small Small Moderate Small Small Moderate Small Small Unreported.
20.	Rota et al. (2016)	N = 200 (102 women); *Age:* 18-35 Gen. pop. (Caucasians).	56 moral dilemmas: - 28 incidental dilemmas (where the sacrifice of one person to save more people is a foreseen but unintended side-effect of action) - 28 instrumental dilemmas (where the sacrifice of one person is the mean to save more people)	MJ (appropriateness) & MDM	IRI (EC, PD, PT, FS)	Chi-square tests (responses in all dilemmas)	**MDM:** FS – UC ($\chi^2_{(1)}$ = 14.6; p <.001): positive association (more FS, more UC) EC – UC ($\chi^2_{(1)}$ = 9.46; p <.001): negative association (more EC, less UC) PT – UC & PD – UC: unreported. **MJ:** UJ – EC: $\chi^2_{(1)}$ = 20.1; p <.001): negative association (the less EC, the more appropriate) → especially for women, not men. UJ – PT/FS/PD: unreported.	Unreported. Unreported. Unreported.
21.	Santamaría-García et al. (2017)	N = 1109 (567 women); M_{age} = 37.6; SD = 12.5 Gen. pop. (workers from Latin American countries).	2 trolley problems: 1 impersonal (Standard fumes) and 1 personal (Footbridge)	MDM	IRI (EC)	Multiple linear regression (DV = EC; IVs = Deontological choice [1 = DC; 0 = UC])	Personal dilemma: $F_{(1,372)}$ = 3.99; p <.01; R^2 = 0.01 DC → EC: *Beta* = 1.33; p < .01 Impersonal dilemma: DC → EC: *ns.*	Small η^2 = 0.07
22.	Sarlo et al. (2014)	N = 37 (19 women) M_{age} = 23.7; SD = 1.9 Gen. pop. (undergraduate students).	30 Footbridge-type dilemmas 30 Trolley-type dilemmas 12 additional moral dilemmas	MDM	IRI (EC, PD, PT, FS)	Correlations Two Multiple Stepwise Regression analyses MSR1: DV = Percentage of UC in Footbridge-type dilemmas; IVs = PT, FS, EC, PD. MSR2: DV = Percentage of UC in trolley-type dilemmas; IVs = PT, FS, EC, PD.	Impersonal dilemma: UC – EC: r = -0.154 (*ns.*) UC – PT: r = -0.135 (*ns.*) UC – PD: r = -0.165 (*ns.*) UC – FS: r = -0.073 (*ns.*) Personal dilemma: UC – EC: r = -0.316 (p < .10) UC – PT: r = 0.063 (*ns.*) UC – PD: r = -0.454 (p = .005) UC – FS: r = -0.093 (*ns.*) MSR1: Only PD was a significant predictor of UC (β = -0.45; B = -2.02; SE = .15; p: Unreported) MSR2: no significant predictors were found.	Moderate Unreported.
23.	Takamatsu (2018)	**Study 1**: N = 275 (118 women); M_{age} = 37.7; SD = 11.95 Gen. pop. (Mturk workers, mostly white) **Study 2**: N = 170 (77 women); M_{age} = 37.7; SD = 11.95 Gen. pop. (Mturk workers, mostly white)	**Study 1**: Footbridge (other-beneficial dilemma) and Raftboat (self-beneficial dilemma) **Study 2:** 6 sacrificial moral dilemmas (Footbridge, Crying baby, Vaccine, Raftboat, Modified safari, Sophie's choice)	MDM (prediction: *Would you push …?*)	**Study 1:** Other-focused empathy (OFE): measured by five adjectives (sympathetic, compassionate, concerned, empathic, tender) Self-focused empathy (SFE): measured by five adjectives (low-spirited, heavy-hearted, sad, sorrowful, melancholy) **Study 2:** EC for the victim (ECv) vs. EC for the saved (ECs) IRI (EC)	**Study 1:** Logistic regressions (DV = UC; IVs= self and other-focused empathy; covariate = gender). **Study 2:** Correlations Logistic regressions (DV = UC; IVs= self and other-focused empathy; covariate = gender). Mediation analysis	**Study 1:** Footbridge dilemma: OFE → UC: b = -.073; *Wald* = 11.96; p = .001 SFE → UC: b = -.095; *Wald* = 20.02; p < .001 Raftboat dilemma: OFE → UC: b = .081; *Wald* = 17.97; p < .001 SFE → UC: b = .068 *Wald* = 10.68; p = .001 **Study 2:** UC – EC: r = -0.14 (*ns.*) UC – EfS: r = -0.083 (*ns.*) UC – EfV: r = -0.22 (p < .01) Footbridge dilemma: EfV → UC: b = -.19; *Wald* = 11.0; p = .001 EfS → UC: b = .14; *Wald* = 5.47; p = .019 Raftboat dilemma: EfV → UC: b = -.13; *Wald* = 4.89; p = .027 EfS → UC: b = .12; *Wald* = 4.07; p = .044 EfS and EfV fully mediated the association between psychopathy and UC. The direct effect of psychopathy was no longer significant after mediating variables were included in the model (p = .58). EC was not a significant mediator.	Unreported. Unreported. Unreported. Unreported. Small Unreported. Unreported. Unreported. Unreported.
24.	Takamatsu & Takai (2019)	N = 282 (223 women); M_{age} = 19.4; SD = .94 Gen. pop. (undergraduate students).	4 personal moral dilemmas: - Crying baby, - Modified lifeboat, - Sophie's choice, - Footbridge.	MJ (appropriateness)	IRI (EC)	Correlations Hierarchical Regression Analysis (DV = UJ; IVs = [Block 1] Age, Gender; [Block 2] Primary psychopathy, Secondary psychopathy, Alexithymia domains; [Block 3] EC)	EC – UJ: r = -0.29 (p < .001) HRA: EC→ UJ: β = -.20; p < .01	Small/Mod. ΔR^2 = .036

(Continues)

TABLE 1 (Continued)

#	Authors and date	Population	Trolley types or measurements	MJ, MI, or MDM	Empathy measurements	Statistics	Main results	Effect sizes
				Records identified via other methods				
25.	Caviola et al. (2021)	**Study 1**: N = 905 (445 women); M_{age} = 41.3; SD = 12.5 Gen. pop. (American Mturk workers). **Study 2**: N = 400 (194 women); M_{age} = 38.4; SD = 11.5 Gen. pop. (American Mturk workers). **Study 3**: N = 113 (45 women); M_{age} = 38.2; SD = 10.9 Gen. pop. (Mturk workers). **Study 7**: N = 301 (143 women); M_{age} = 41.7; SD = 12.9 Gen. pop. (American Mturk workers). **Study 8:** N = 233 (94 women); M_{age} = 37.7; SD = 11 Gen. pop. (American Mturk workers). **Study 10:** N = 201 (93 women); M_{age} = 40.9; SD = 12.9 Gen. pop. (American Mturk workers).	Oxford Utilitarianism Scale: - Permissive attitude toward instrumental harm (OUS-IH) - Impartial concern for the greater good (OUS-IB) 10 trolley problem variants (animals vs. humans): incongruent vs. congruent dilemmas (Conway and Gawronski, 2013)	Moral Inclination & MJ (permissibility/appropriateness of the utilitarian action)	IRI (EC)	Correlations	**Study 1: (humans and pigs)** EC – OUS-IB: r = 0.37 (p < .001) EC – OUS-IH: r = -0.29 (p < .001) Permissibility – EC: r = -0.16 (p < .001) Harm aversion – EC: r = 0.32 (p < .001) **Study 2: (humans and zebras)** EC – OUS-IB: r = 0.47 (p < .001) EC – OUS-IH: r = -0.15 (p < .001) Permissibility – EC: r = -0.09 (*ns.*) Harm aversion – EC: r = 0.37 (p < .001) **Study 3: (humans and pigs)** EC – OUS-IB: r = 0.29 (p < .01) EC – OUS-IH: r = -0.15 (p < .05) EC - UP: r = 0.01 (*ns.*) EC - DP: r = 0.15 (*ns.*) Harm aversion – EC: r = 0.49 (p < .001) Humans: - EC - UP: r = -0.02 (*ns.*) - EC - DP: r = 0.23 (*ns.*) Animals (pigs): - EC - UP: r = 0.03 (*ns.*) - EC - DP: r = 0.13 (*ns.*) **Study 7: (humans and dogs)** Permissibility – EC: r = -0.07 (*ns.*) Harm aversion – EC: r = 0.25 (p < .001) **Study 8: (humans with cognitive impairments and chimpanzees)** Permissibility – EC: r = -0.25 (p < .001) Harm aversion – EC: r = 0.26 (p < .001) **Study 10: (humans with a persistent vegetative state and puppies)** Permissibility – EC: r = -0.17 (p < .01) Harm aversion – EC: r = 0.44 (p < .001)	Moderate Small/Mod. Small Moderate Moderate Small Moderate Small/Mod. Small Moderate Small Small Small Moderate
26.	Cecchetto et al. (2018)	N = 43 (21 women) M_{age} = 24.7; SD = 3.69 Gen. pop. (native Italian speakers).	48 dilemmas based on four conceptual factors (personal force, Intentionality, beneficient recipient, and evitability)	MDM	IRI (EC, PD, PT, FS)	Linear Mixed-Effects Model DV = UC; IVs = Gender, Affective and Cognitive components of the Bermond-Vorst Alexithymia Questionnaire (Bermond and Oosterveld, 1994), Toronto Alexithymia scale total score (Bagby et al., 1994), all IRI domains, Personal force, Benefit recipient, Intentionality) Spearman's correlations (Total utilitarian score on the 48 dilemmas [UC=1])	PT → UC: β = -.041; *ns.* PD → UC: β = .027; *ns.* EC → UC: β = -.063; *ns.* FS → UC: β = .042; *ns.* UC – PT: r = -0.16 (*ns.*) UC – PD: r = -0.103 (*ns.*) UC – EC: r = -0.174 (*ns.*) UC – FS: r = -0.157 (*ns.*)	
27.	Conway & Gawronski (2013)	**Study 1**: N = 112 (82 women); M_{age} = 19.2; SD = 5.2 Gen. pop. (undergraduates, mostl Caucasians and Christians). **Study 3:** N = 275 (156 women); M_{age} = 34.1; SD = 11.7 Gen. pop. (American Mturk workers, mostly Caucasians).	10 trolley problem variants: incongruent vs. congruent dilemmas (Conway and Gawronski, 2013)	MJ (appropriateness of Deontological action) & Moral Inclination	IRI (EC & PT) EC manipulation	Correlations Multiple Linear Regression Analyses (DV = UP/DP; IVs = EC, PT, Faith in intuition, Need for cognition, Religiosity, Moral identity internalization) Mixed Model ANOVA (Parameter: Deontology vs. utilitarianism X Processing: EC vs. Control condition)	**Study 1:** MJ – EC: r = 0.23 (p < .01) DP – EC: r = 0.28 (p < .01) DP – PT: r = 0.32 (p < .01) MJ – PT: r = 0.31 (p < .01) UP – EC: r = -0.01 (*ns.*) UP – PT: r = -0.12 (*ns.*) **Study 3:** MLR: EC → DP: β = 0.28; p < .01 PT → DP: β = 0.34; p < .01 EC → UP: β = -0.04; p = .71 PT → UP: β = -0.15; p = .10 MM ANOVA: deontological inclinations were significantly higher in the EC condition than in the control condition: $F_{(1,273)}$ = 6.40; p = .012 Utilitarian inclinations were unaffected by the EC manipulation: $F_{(1,273)}$ = 0.39; p = .534	Small Small/Mod. Moderate Moderate η_p^2 = 0.023
28.	Nasello et al. (2023)	N = 427 (231 women); M_{age} = 25.3; SD = 5.10 Gen. pop. (graduates and mostly Caucasians	1 transplant dilemma (3 perspectives) 1 everyday trolley-like problem (3 perspectives) *Perspective 1:* the participant is the bystander and has to decide between sacrificing the one to save the five or inversely. *Perspective 2:* the participant has to choose between sacrificing him/herself to save the five or inversely. *Perspective 3:* the participant is among the five and has to decide between sacrificing the one to save the five (including him/her) or inversely.	MDM	BES (affective & cognitive empathy [AE, CE])	Correlations General Mixed Model analyses (DV = UC; IVs = perspectives, AE, CE, borderline traits)	Transplant: AE – UC (P1): r = -0.09; *ns.* AE – UC (P2): r = -0.03; *ns.* AE – UC (P3): r = -0.16; p < .01. CE – UC (P1): r = -0.11; p < .05. CE – UC (P2): r = -0.08; *ns.* CE – UC (P3): r = -0.16; p < .01. Everyday TLP: AE – UC (P1): r = 0.05; *ns.* AE – UC (P2): r = 0.15; p < .01 AE – UC (P3): r = 0.04; *ns.* CE – UC (P1): r = 0.03; *ns.* CE – UC (P2): r = 0.13; p < .01. CE – UC (P3): r = 0.03; *ns.* GMM: Transplant: AE → UC: B = -0.007; SE = 0.025; p = .776 CE → UC: B = -0.062; SE = 0.038; p = .102 Everyday TLP: AE → UC: B = 0.037; SE = 0.014; p = .006 CE → UC: B = -0.013; SE = 0.022; p = .567	Small Small Small Small Small Exp(B) = 1.1
29.	Park et al. (2016)	N = 64 (38 women) M_{age} = 19.4 Gen. pop (undergraduate student mostly Caucasians).	10 trolley problem variants: incongruent vs. congruent dilemmas (Conway and Gawronski, 2013)	MJ (inappropriateness of the utilitarian action*) & Moral Inclination *higher values indicate stronger deontological tendencies.	IRI (EC & PT)	Correlations Multiple Linear Regression (DV = UP/DP; IVs = EC, Need for cognition, the Duke University Religion Index, Heart Rate Variability)	EC – UJ: r = .23 (*ns.*) DP – EC: r = 0.37 (p < .05) DP – PT: r = 0.09 (*ns.*) PT – UJ: r = .13 (*ns.*) UP – EC: r = 0.05 (*ns.*) UP – PT: r = -0.12 (*ns.*) MLR: EC → DP: β = 0.37; p = .01 EC → UP: β = 0.05; p = .72	Moderate Unreported.
30.	Patil et al. (2016)	*ASD participants:* N = 15 (6 women); M_{age} = 37.4; SD = 13.02; education year: M = 3.4 (1.92) *HC:* N = 17 (4 women); M_{age} = 32.03; SD = 9.44; education year: M = 4.5 (1.41)	18 dilemmas: - 6 non-moral dilemmas - 6 personal moral dilemmas - 6 impersonal moral dilemmas	MDM (*Would you do it?*)	IRI (EC, PD)	Path analysis in ASD DV = UC in personal moral dilemmas; Regressors: EC, PD, trait alexithymia, trait autism, medication status, non-verbal intelligence. Correlations	EC → UC: β = -0.491; SE = 0.132; p = .001 PD → UC: β = -0.42; SE = 0.157; p = .013 Impersonal dilemmas (M_{ASD-HC}) EC – UC: r = -0.252 (*ns.*) PT – UC: r = 0.014 (*ns.*) PD – UC: r = -0.0595 (*ns.*) Personal dilemmas (M_{ASD-HC}) EC – UC: r = -0.269 (*ns.*) PT – UC: r = -0.0585 (*ns.*) PD – UC: r = -0.254 (*ns.*)	Unreported. Unreported.

TABLE 1 (Continued)

#	Authors and date	Population	Trolley types or measurements	MJ, MI, or MDM	Empathy measurements	Statistics	Main results	Effect sizes
31.	Reynolds and Conway (2018)	**Study 2:** $N = 296$ (136 women); $M_{age} = 35.8$; $SD = 12.8$ Gen. pop. (American Mturk workers).	10 trolley problem variants: incongruent vs. congruent dilemmas (Conway and Gawronski, 2013)	MJ (appropriateness) & Moral Inclination	IRI (EC)	Correlations Multiple Linear regressions DV = EC; IVs = UP, DP (parameters not completely described)	**Study 2:** EC – UJ: $r = -.23$; $p < .001$ EC – DP: $r = .30$; $p < .001$ EC – UP: $r = .02$; *ns.* MLR: UP → EC: $\beta = -.01$; $p = .833$. DP → EC: : $\beta = .30$; $p < .001$	Small Moderate Unreported.
32.	Romero-Rivas et al. (2022)	$N = 294$ (273 women); $M_{age} = 19.3$; $SD = 1.70$ Gen. pop. (education students).	2 trolley problems: 1 impersonal (Standard fumes) and 1 personal (Footbridge)	MDM *What should you do? (multiple choices)* - *Nothing* - *Push the stranger/Flip to the right* - *Jump onto the track/Flip to the left*	BES (affective and cognitive empathy)	Correlations (UC was coded 1, other choices 0) Bayesian one-way ANOVA (testing whether the empathy scores varied depending on the participant's responses to the dilemmas)	AE – UC: $r = -0.037$ (*ns.*) CE – UC: $r = 0.026$ (*ns.*) Bayesian ANOVA: No significant differences: neither in affective empathy nor cognitive empathy ($Fs < 1$; $ps > .87$; $BFs < .07$)	
33.	Takamatsu (2019)	$N = 439$ (206 women); $M_{age} = 35.5$; $SD = 11.5$ Gen. pop. (American Mturk workers, mostly Caucasians).	2 everyday trolley-like problems (Vyas et al., 2017): ostracism & job termination	MDM & MJ	Empathy for the target; items adapted from (Hepper et al., 2014): Affective empathy (AE) Cognitive empathy (CE)	Ordinal Regression Analysis (DV = UC/UJ; IVs = Primary psychopathy, social dominance orientation, dehumanization beliefs, AE, CE; controlling gender) Correlations	ORA (Utilitarian Judgments): <u>Ostracism</u>: AE → UJ: $\beta = -.16$; *Wald* = 20.8; $p < .001$ CE → UJ: $\beta = -.07$; *Wald* = 3.26; *ns.* AE – UJ: $r = -.33$; $p < .001$ CE – UJ: $r = -.36$; $p < .001$ <u>Job</u>: AE → UJ: $\beta = -.15$; *Wald* = 14.1; $p < .001$ CE → UJ: $\beta = -.03$; *Wald* = 1.14; *ns.* AE – UJ: $r = -.33$; $p < .001$ CE – UJ: $r = -.37$; $p < .001$ ORA (Utilitarian Choices): <u>Ostracism</u>: AE → UC: $\beta = -.18$; *Wald* = 12.2; $p < .01$ CE → UC: $\beta = -.10$; *Wald* = 4.34; $p < .05$ AE – UC: $r = -.29$; $p < .001$ CE – UC: $r = -.31$; $p < .001$ <u>Job</u>: AE → UC: $\beta = -.06$; *Wald* = 1.28; *ns.* CE → UC: $\beta = -.10$; *Wald* = 4.05; $p < .05$ AE – UC: $r = -.27$; $p < .001$ CE – UC: $r = -.25$; $p < .001$	Odds: 0.83 Odds: 0.85 Odds: 0.84 Odds: 0.90 Small/Mod. Moderate Odds: 0.90 Small Small
34.	Zhang et al. (2020)	$N = 310$ (234 women); $M_{age} = 19.72$; $SD = 2.48$ Gen. pop. (undergraduate students).	10 trolley problem variants: incongruent vs. congruent dilemmas (Conway and Gawronski, 2013)	Moral inclination & MJ (acceptableness)	IRI (EC)	Correlations Mediation analyses	EC – UJ: $r = -0.22$; $p < .05$ EC – UP: $r = 0.063$; *ns.* EC – DP: $r = 0.356$; $p < .01$ Mediation analyses: EC → DP: $\beta = .25$; $p = .03$ EC → UJ: $\beta = .002$; *ns.* EC → UP: $\beta = .05$; *ns.* EC → UJ: $\beta = -.19$; $p < .05$	Small Moderate Unreported. Unreported.

Note: This table describes all retained records in detail.

Abbreviations: AD, Affective Dissonance; AE, Affective Empathy; AR, Affective Resonance; ASD, Autism Spectrum Disorders; BES, Basic Empathy Scale; CA, Consequentialism Appropriateness; CE, Cognitive Empathy; DA, Deontology Appropriateness; DC, Deontological Choice; DJ, Deontological Judgements; DP, Deontology Parameter; DV, Dependent Variable; EC, Empathic Concern; ECs, Empathic Concern for the Saved; ECv, Empathic Concern for the Victim; EQ, Empathy Quotient; Everyday TLP, Everyday Trolley-Like Problem; FAB, Frontal Assessment Battery; FS, Fantasy; Gen. pop., General population; GMM, General Mixed Model analysis; HC, Healthy Controls; HRA, Hierarchical Regression Analysis; IRI, Interpersonal Reactivity Index; IV, Independent Variable; KP, Korsakoff patients; LMRM, Linear Mixed Regression Model; MDM, Moral Decision-Making, MLR, Multiple Linear Regression; MJ, Moral Judgements; MSR, Multiple Stepwise Regression; ns., non-significant; OFE, Other-Focused Empathy; ORA, Ordinal Regression Analysis; OUS-IB, Oxford Utilitarian Scale, Impartial concern for the greater good; OUS-IH, Oxford Utilitarian Scale, Permissive attitude toward instrumental harm; PD, Personal Distress; PF1, PF2, Personality Factor 1, 2; PT, Perspective-Taking; SFE, Self-Focused Empathy; UC, Utilitarian Choice; UJ, Utilitarian Judgement; UMR, Utilitarian Moral Reasoning; UP, Utilitarian Parameter; U-responders, Utilitarian responders.

All meta-analyses were conducted using the Jamovi software, version 2.3.21 (The Jamovi Project, 2019). We performed *random* effects models on all meta-analyses and measured the estimate of each model, the heterogeneity, and the publication bias. In addition, we reported the forest and funnel plots for each meta-analysis. The model estimator was a *restricted maximum-likelihood*, the model measures were *Fisher's r-to-z transformed correlation coefficient*, and we applied the *Knapp and Hartung adjustment* (when necessary). In total, six meta-analyses were conducted: two studying the estimated correlation between empathy domains and moral judgements (between affective empathy and utilitarian judgements and between cognitive empathy and utilitarian judgements), two studying the estimated correlation between empathy domains and moral decision-making (between affective empathy and utilitarian choices, and between cognitive empathy and utilitarian choices), and two studying the estimated correlation between empathic concern and moral inclination (between EC and the utilitarian and deontological parameters).

RESULTS

Descriptive information

The 34 reviewed studies were published between 2012 and 2023. The average sample size for studies using general population samples (e.g., students, Mturk workers, adults) was 594 participants per study

(*General Mean*$_{age}$ = 34.94; N_{total} = 26,015), while the average sample size for studies using specific populations (e.g., clinical populations or experts in moral philosophy) was 69 participants per study, including healthy controls (*General Mean*$_{age}$ = 40.6; N_{total} = 383). Although the records included studies from around the world (e.g., Japan, South America, America, or Europe), most participants were Caucasian.

Twenty-seven studies[1–3,5,7–13,16–27,29–31,34] used the Interpersonal Reactivity Index (IRI, Davis, 1980, 1983), and some focused on specific IRI domains (see Table 1), three used the Basic Empathy Scale (BES, Jolliffe & Farrington, 2006)[14,28,32], one used the Affective and Cognitive Measure of Empathy (Vachon & Lynam, 2016)[6], one used the Empathy Quotient (EQ, Baron-Cohen & Wheelwright, 2004)[8], and four used empathy-related adjectives to assess empathy[4,15,23,33]. Thirteen studies focused on moral decision-making[1,4,6–8,14,21–23,26,28,30,32], seven focused on moral judgements[2,9,15,17–19,24], four focused on moral judgements and moral decision-making[3,11,20,33], three focused on moral inclination[10,13,16], and seven focused on moral inclination and moral judgements[5,12,25,27,29,31,34]. Finally, all of the studies presented correlational designs, except for two studies that utilized some perspective-taking or empathic concern inductions, evidencing the causal effects of empathy on moral decision-making[4] and moral inclination[27]. Nonetheless, it is worth noting that several studies presented confusion between moral judgements and moral decision-making[1,3,4,7,21,30]; these studies were reclassified according to the definitions presented in the introduction.

Utilitarian judgements and empathy domains

Meta-analyses of the associations between affective and cognitive empathy and utilitarian judgements

The first meta-analysis was conducted to study the correlation between affective empathy and utilitarian judgements (judging the appropriateness of the utilitarian action). This meta-analysis contained $k = 27$ studies and showed a significant estimated association ($E = -0.198$; $SE = 0.015$; $p < .001$; $CI^- = -0.230$; $CI^+ = -0.167$; see Figure 2a displaying the Forest plot). There was significant heterogeneity in the size of this association across the included records, $Q_{(26)} = 80.8$; $p < .001$; $\tau^2 = 0.004$; $I^2 = 59.4\%$. We analysed the externally standardized residuals (see Figure S1a) and found that two studies (Hannikainen et al., 2018; Takamatsu, 2019) had larger values than the rest of the sample. However, these studies were not considered outlying studies, as none of the studies had studentized residuals larger than |3.113|. Removing these two studies from the random effects model indicated that they caused the heterogeneity of the model (the heterogeneity decreased to 33.3%, $p = .113$). Hence, although there is some heterogeneity in the $k = 27$ model, the true outcomes of the studies are consistent with the estimated average outcome (100% of the estimates were negative). In addition, removing the two marginal studies (Hannikainen et al., 2018; Takamatsu, 2019) did not affect the significance of the model ($E = -0.180$; $p < .001$), indicating the $k = 27$ random effects model could be maintained (but we applied the Knapp and Hartung adjustment). Concerning publication bias, neither Beff and Mazumdar's rank correlation nor Egger's regression indicates any funnel plot asymmetry (−0.077; $p = .591$; 0.372; $p = .713$, respectively), and the fail-safe number was 3504 ($p < .001$), indicating no publication bias (see also Figure S1b,c displaying funnel and Q–Q plots).

The second meta-analysis investigated the correlation between cognitive empathy and utilitarian judgements (judging the appropriateness of the utilitarian action). The meta-analysis included $k = 9$ studies, but a high level of heterogeneity was found in the size of this association across the included records, $Q_{(8)} = 43.96$; $p < .001$; $\tau^2 = 0.014$; $I^2 = 86.8\%$. The model identified Takamatsu (2019) as an outlier (the studentized residual value was larger than |2.77|, see also Figure S2a). After conducting the analysis without this study, we found that Takamatsu (2019) caused this high level of heterogeneity, and the heterogeneity decreased to 37.7% ($p = .096$) after removing it. However, none of the studies were considered overly influential according to Cook's distances, and neither the Beff and Mazumdar rank correlation nor Egger's regression displayed any asymmetry in the funnel plot ($p = .612$; $p = .831$, respectively). Given these factors and the fact that removing Takamatsu (2019) did not affect the overall estimate's significance, we maintained the $k = 9$ model and

applied the Knapp and Hartung adjustment. ($E = -0.141$; $SE = 0.04$; $p < .05$; $CI^- = -.243$; $CI^+ = -.039$; see Figure 2b displaying the Forest plot).[3] The result showed a significant negative association between cognitive empathy and utilitarian judgements. Finally, the fail-safe number was 230 ($p < .001$), indicating no publication bias (see Figure S2b,c displaying funnel and Q–Q plots).

A systematic review of the associations between empathy domains and utilitarian judgements

Eighteen of the 34 reviewed papers focused on moral judgments[2,3,5,9,11,12,15,17–20,24,25,27,29,31,33,34]. In those papers, the authors investigated how participants assessed the appropriateness or the wrongness of utilitarian or deontological actions in trolley problems and variants. Among these reviewed papers focusing on moral judgements, 15 used the IRI[2,3,5,9,11,12,17–20,24,25,27,31,34] (only one study[2] created a new measurement based on the IRI and other personality measurements[4]), and two used empathy adjectives[15,33] to study the associations between empathy and moral judgements.

Most IRI studies showed that EC was significantly associated (in a negative way) with utilitarian judgments[3,5,9,11,12,17,19,20,24,25,27,31,34], especially in personal moral dilemmas[3,9,11,17,19,24]. However, some of these significant associations disappeared[11,17,34] or decreased[24] in regression or mediation analyses (e.g., EC had no significant direct effect on utilitarian judgements when the Deontological parameter was the dependent variable, but it had a significant direct effect on utilitarian judgements when the Utilitarian parameter was the dependent variable[34]; see Table 1). Overall, the correlations appeared small[3,5,11,12,17,19,25,27,31,34] or moderate[9,24] when significant, and the relation between EC and moral judgements (judging the appropriateness of the utilitarian action) was negative in most cases: the more EC, the less appropriate the utilitarian action. One reviewed paper[25] performed ten studies to assess the variations of utilitarian judgements using trolley problems involving humans and animals. They consistently found small associations between EC and permissibility of the utilitarian action (Mean $r = -.15$; see Table 1). In rare cases, the significant relation occurred only in women but not men[20] and was found to be positive in specific scenarios[19] (i.e., *sacrifice comfort* and *let mother die*[5]). However, following a cluster analysis[19], some authors found a contrasting result showing that participants who considered utilitarian actions appropriate in all scenarios had higher EC levels than those who rated utilitarian actions as wrong. Only a few studies have found no significant relationship between EC and utilitarian judgments[29] in impersonal[12,17,18] and personal[18] moral dilemmas: one of these studies included both anxiety disorder patients and healthy controls[18], while another included trolley scenarios involving animals[25]. Additionally, other authors[25] reported insignificant associations between EC and permissibility (see their studies 2 and 7; see Table 1).

Concerning other IRI domains (PT, PD, and/or FS), seven studies included these domains in their analyses[2,3,9,17,18,20,27]. PT was not significantly related to utilitarian judgements in four studies[3,17,18,20] (some used only personal moral dilemmas[3] or both personal and impersonal scenarios[17,18,20]). Only two studies[9,27] found a significant correlation between PT and utilitarian judgements, $r = -.11$; $r = (-).31$,

[3]The $k = 8$ random effects model (without Takamatsu, 2019) showed an overall estimate = −0.097 ($SE = 0.03$; $p < .01$).

[4]The authors conducted a study where they measured 17 personality traits and used an exploratory factor analysis to identify five new factors. PF1 includes altruism, faith in intuition, victim injustice sensitivity, less callous affect, fantasy, empathic concern, and perspective taking (considered as other-oriented facets of empathy). PF2 is characterized by increased neuroticism, personal distress with decreased self-esteem. PF3 is characterized by increased trait psychopathy, lower altruism, and increased justice sensitivity from the perpetrator's perspective. PF4 includes general justice sensitivity across all perspective (perpetrator, victim, observer, beneficiary). Finally, PF5 is characterized by strong faith in intuition and little motivation for critical, analytical thinking, increased obedience to authority, and greater sensitivity from the victim perspective.

[5]In *let mother die*, a firefighter is faced with the decision of saving either his uneducated housekeeper, who is also his mother, or a highly successful and still active peace negotiator. In *sacrifice comfort*, a person must choose whether to sacrifice some of their own well-earned comfort by donating money to help save people from starving to death in other parts of the world. In both scenarios, individuals who scored higher on EC tended to judge more morally wrong to not donate to charity or to not save the peace negotiator.

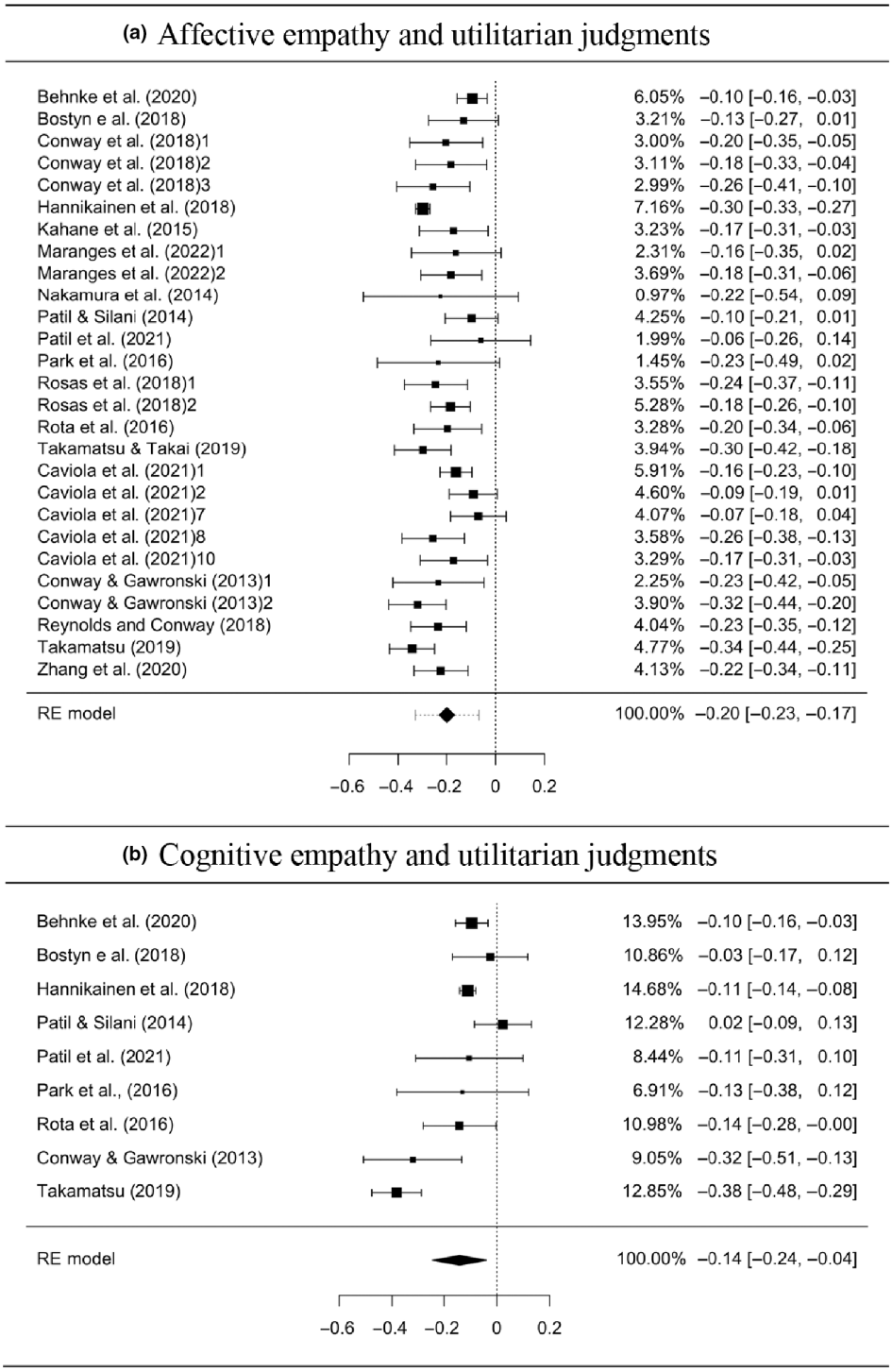


FIGURE 2 Forest plots of the meta-analytic correlations between empathy domains and utilitarian judgements. (a) Affective empathy and utilitarian judgements. (b) Cognitive empathy and utilitarian judgements. Error bars indicate 95% CIs.

respectively; see Table 1, but one of these significant associations disappeared in a hierarchical regression model[9]. Regarding FS, most studies found insignificant correlations with utilitarian judgments[9,17,20], although one record[7,study3] reported a (small) significant positive predictive effect of FS on utilitarian judgements in a hierarchical regression model. Finally, PD was not significantly related to utilitarian judgements in three studies (regardless of using personal or impersonal moral dilemmas)[17,18,20]. However, one study that used personal moral dilemmas reported a small significant correlation between PD and utilitarian judgements ($r = -.08$; see Table 1) and a small significant (negative) predictive effect of PD on utilitarian judgments[9].

A separate analysis showed a significant negative correlation between PF1[2] (a domain including FS, PT, and EC) and utilitarian judgements in a personal moral dilemma (Mean $r = -.095$; see Table 1) and a negative predictive effect of PF1 on utilitarian judgements when using personal scenarios (i.e., transplant and footbridge dilemmas, qualified as other-oriented scenarios by the authors). Overall, the effects of this new factor (including IRI domains) were small when significant. On the other hand, PF2 (a domain including PD) was not significantly associated with utilitarian judgements.

Lastly, two studies used empathy adjectives to study their associations with utilitarian judgements in traditional personal scenarios[15] and everyday trolley-like problems[33]. One study[15] found a significant correlation between empathy and utilitarian judgements in high-conflict scenarios (comparable to personal scenarios, $r = -.424$; see Table 1) and a significant (and negative) predictive effect of empathy on utilitarian judgements. The other study found that affective and cognitive empathy were significantly correlated (in a negative way) with utilitarian judgements in two everyday trolley-like problems (job termination and ostracism)[33]. However, only affective empathy predicted utilitarian judgements in these two dilemmas (not cognitive empathy), but the effect size of this predictor was small (see Table 1).

Utilitarian choices and empathy domains

Meta-analyses of the associations between affective and cognitive empathy and utilitarian choices

We conducted a meta-analysis to investigate the overall association between affective empathy and utilitarian choices. The meta-analysis included $k = 12$ studies and reported a significant overall estimate ($E = -0.151$; $SE = 0.04$; $p = .005$; $CI^- = -.244$; $CI^+ = -.057$; see Figure 3a displaying the Forest plot), showing a negative association between affective empathy and utilitarian choices. However, we found a high level of heterogeneity in the size of this association across the included records, $Q_{(11)} = 55.2$; $p < .001$; $\tau^2 = 0.02$; $I^2 = 80.5\%$. The examination of the studentized residuals revealed that none of the studies had a value larger than |2.865| (see Figure S3a). According to Cook's distances, none of the studies could be considered overly influential, and neither Beff and Mazumdar's rank correlation nor Egger's regression indicated any asymmetry in the funnel plot ($p = .459$; $p = .639$, respectively). Hence, no outlier was identified, so we investigated whether age or empathy measurements contributed to this high heterogeneity. We found that age was a significant moderator ($E = -0.01$; $SE = 0.01$; $p < .05$), but not empathy measurements ($p = .593$), which partly explained the heterogeneity of the random effects model (see Tables S1–S4). Considering these factors, we applied the Knapp and Hartung adjustment in the $k = 12$ model. Finally, the fail-safe number was 254 ($p < .001$), indicating no publication bias (see also Figure S3b,c displaying funnel and Q–Q plots).

Another meta-analysis investigated the association between cognitive empathy and utilitarian choices, using data from nine studies ($k = 9$). The analysis showed an insignificant overall estimate ($E = -0.07$; $SE = 0.04$; $p = .075$; $CI^- = -.154$; $CI^+ = .009$; see Figure 3b displaying the Forest plot), showing no association between cognitive empathy and utilitarian choices. However, the Q-Test indicated a significant amount of heterogeneity in the true outcomes, $Q_{(8)} = 26.3$; $p < .001$; $\tau^2 = 0.009$; $I^2 = 68.2\%$.

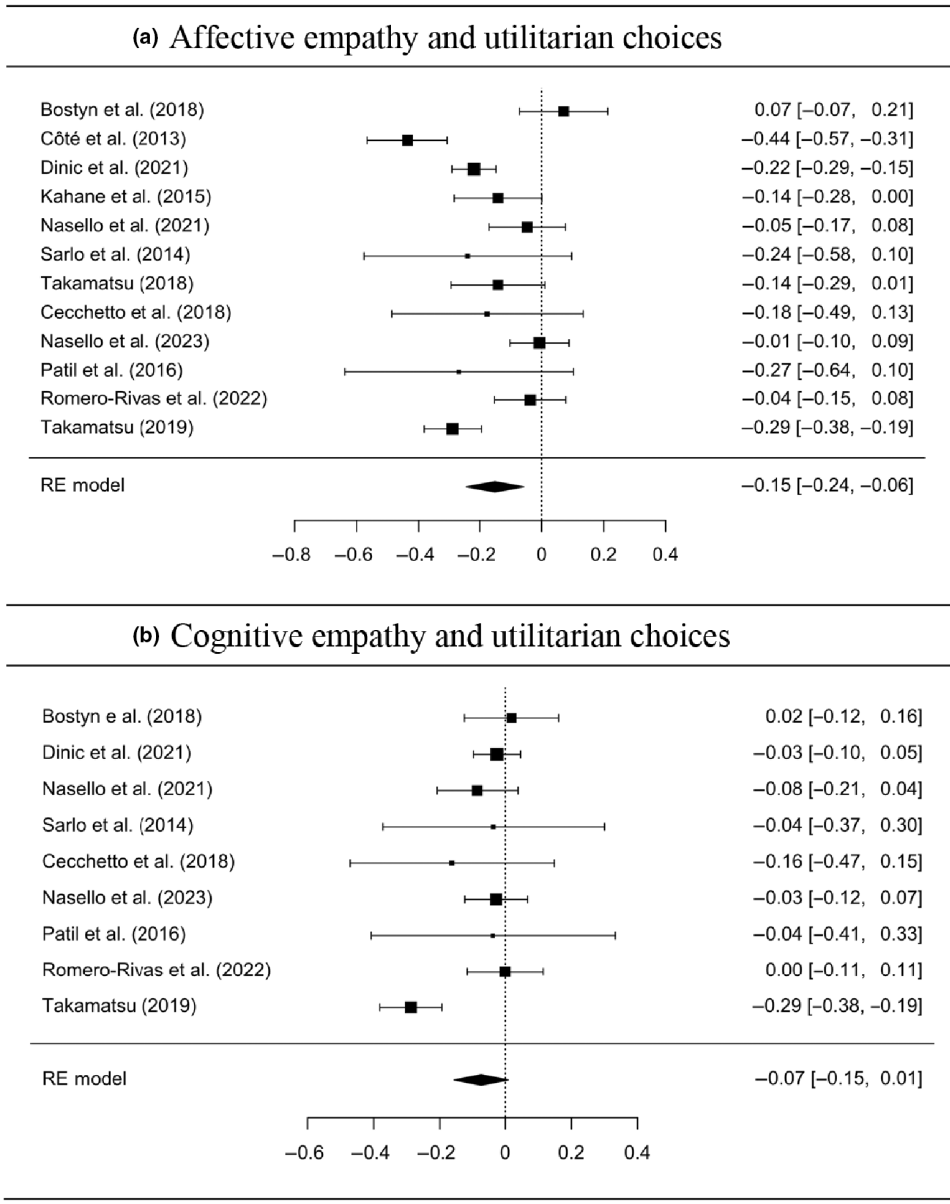


FIGURE 3 Forest plots of the meta-analytic correlations between empathy domains and utilitarian choices. (a) Affective empathy and utilitarian choices. (b) Cognitive empathy and utilitarian choices. Error bars indicate 95% CIs.

We found that Takamatsu (2019) appeared as an outlier in the model, with a studentized residual value larger than |2.773| (see also Figure S4a). After removing this study from the analysis, we found that Takamatsu (2019) caused this high level of heterogeneity (after removing Takamatsu, 2019: $Q_{(17)} = 2.179$; $p = .949$; $\tau^2 = 0.002$; $I^2 = 0\%$), but removing this study did not affect the significance of the $k = 9$ model (which prompted us to maintain the $k = 9$ model and apply the Knapp and Hartung adjustment). Neither Beff and Mazumdar's rank correlation nor Egger's regression displayed any funnel plot asymmetry ($p = .584$; $p = .657$, respectively; see Figure S4b,c displaying funnel and Q–Q plots). The fail-safe number was equal to 27 ($p < .001$), indicating caution in interpreting the results of this meta-analysis, but removing the outlier (Takamatsu, 2019) reduced the fail-safe number to zero ($p = .092$).

A systematic review of the associations between empathy domains and utilitarian choices

Seventeen reviewed papers studied the links between empathy and utilitarian choices[1,3,4,6–8,10,14,20–23,26,28,30,32,33]. EC (the IRI domain) was the most selected empathy component from these reviewed studies.

These reviewed papers can be divided into two categories. On the one hand, some studies[3,11,22,23,26,30] showed no significant relationship between EC and utilitarian choices (one study[3] used a real-life trolley-type problem) or insignificant differences in EC scores for those who made deontological choices[8]. On the other hand, other studies revealed some significant (negative) associations or predictive effects between these variables[1,20,21,30] (a study[30] displayed a significant predictive effect but no significant correlation), or higher EC scores for those who made deontological choices[7]. These negative associations are depicted as follows: the more EC, the fewer utilitarian choices (thus, the more deontological choices). The significant links or differences occurred mostly in personal moral dilemmas[1,7,21,30], but some authors found this relation in both personal and impersonal moral dilemmas[20]. Interestingly, two of these studies conducted multiple regression analyses by setting empathy domains as dependent variables and utilitarian choices as an independent variable[1,21]; they showed that utilitarian/deontological choices significantly predicted EC in personal moral dilemmas (see Table 1). Finally, a discriminant analysis[7] revealed that EC was the factor that most strongly classified utilitarian participants apart (among factors like age, gender, education, scores in the Moral Behaviour Inventory [Mendez et al., 2005], scores in the Daily Spiritual Experience Scale [Underwood & Teresi, 2002], and the other three IRI domains). In conclusion, the studies that reported effect sizes suggest that the association between EC and utilitarian/deontological choices is small[1] or moderate[7,21].

The reviewed studies that examined the links between the other IRI domains (PT, FS, and PD) and utilitarian choices[1,3,7,8,20,22,26,30] showed that: (1) PD was not associated with utilitarian choices in most cases[1,20,26] and utilitarian and deontological responders showed similar PD levels in both personal and impersonal moral dilemmas[7,8]. Only two studies reported that PD was a significant negative predictor of utilitarian choices in personal moral dilemmas[22,30]. However, it is important to note that these studies have small sample sizes (<40 participants): one was conducted on participants diagnosed with autism spectrum and showed no significant correlations between utilitarian choices and PD[30] (contrasting the significant predictive effect); the other showed a significant correlation only in personal moral dilemmas[22]. (2) Most studies reported an insignificant association between FS and utilitarian choices[1,22,26] or insignificant differences in FS levels between utilitarian and deontological responders in personal and impersonal moral dilemmas[7]. Only one study[20] found a significant association between FS and utilitarian choices in personal and impersonal moral dilemmas (but the magnitude of this association was not reported). (3) Most studies showed no significant associations between PT and utilitarian choices[1,3,20,22,26] or that the utilitarian and deontological responders differed in their PT levels[7]. Only one small significant difference between utilitarian and deontological responders was found in personal moral dilemmas, but not in impersonal dilemmas[8] (in a population of participants presenting autistic traits): deontological responders presented significantly higher PT scores than utilitarian responders. Finally, when using a PT induction methodology[4] (for the losing member) or a real-life trolley-type dilemma[3], PT still does not significantly predict utilitarian choices, confirming that PT is not associated with utilitarian choices.

Among the reviewed papers, several authors used alternative measures of empathy (such as the BES [Jolliffe & Farrington, 2006] or empathy-related adjectives). For example, among the three studies[14,28,32] that used the BES, two showed significant associations between affective or cognitive empathy [AE, CE] and utilitarian choices under specific conditions[14,28]. The significant associations between empathy domains and utilitarian choices depended on gender, the type of trolley problem, and the participant's perspective. The authors used some everyday trolley-like problems (i.e., some *causing inconvenience* scenarios) and a personal moral dilemma (the transplant). One particularity of these studies was to invite the participant to be the bystander, the isolated person, or the one among the five (see Table 1). The

authors[28] found that: (1) AE and CE were significantly associated with utilitarian choices when the participant was the isolated person in the everyday trolley-like problem[6] ($r = .15/.13$, see Table 1), (2) AE and CE were significantly associated with utilitarian choices when they were the one among the five in the transplant dilemma ($r = -.16$, see Table 1), and (3) that only CE was significantly correlated with utilitarian choices ($r = -.11$, see Table 1) when they were the doctor (i.e., the bystander) in the transplant dilemma. However, when performing regression models, only AE significantly predicted utilitarian choices (but the effect size was small) in the everyday trolley-like problem (in the perspective where the participant had to decide between saving their job or their five colleagues' jobs). The authors found that the more AE, the more utilitarian choices. In their other study using another everyday trolley-like problem[14], (1) the more AE, the less utilitarian choices when male participants were the bystander, and (2) men and women had different types of choices at low AE levels when they were the isolated person (i.e., males made less utilitarian choices and women made more utilitarian choices). However, as in study[28], the overall predictive effect of AE on utilitarian choices was small (see Table 1). Finally, the last study that used the BES found no significant association between empathy domains and utilitarian choices[32].

One study[6] used the affective and cognitive measure of empathy (Vachon & Lynam, 2016) and found significant associations between affective resonance and affective dissonance (two empathy domains) and utilitarian choices in both personal and impersonal moral dilemmas (with stronger associations in personal scenarios). However, only affective dissonance had a significant negative effect on predicting utilitarian choices. The overall variance explained by all empathy domains was 4% in impersonal scenarios and 9% in personal scenarios.

When authors used empathy adjectives to measure empathy[4,23,33], one author found a small but significant predictive effect of AE or CE on utilitarian choices[33]. In an ostracism everyday trolley-like problem, AE and CE negatively predicted utilitarian choices, but only CE negatively predicted utilitarian choices in a job termination everyday trolley-like problem. Interestingly, another study[23] subdivided *empathy concern for the victim* and *empathy concern for the saved* and found a small negative correlation between empathic concern for the saved and utilitarian choices ($r = -.22$, see Table 1). In addition, they showed that in the footbridge and raftboat[7] dilemmas, both empathic concern for the saved and the victim significantly predicted utilitarian choices, but in opposite ways (positively for the former, negatively for the latter; see Table 1). Finally, they subdivided empathy into *other-focused* (i.e., empathic concern) and *self-focused* indices (i.e., personal distress). They found that both other and self-focused empathy negatively predicted utilitarian choices in the footbridge dilemma and positively predicted utilitarian choices in the raftboat dilemma (see Table 1). Lastly, other authors[4] showed that empathy negatively predicted utilitarian choices in an allocation money task. In this study, empathy explained an important part of the variance in comparison to other parameters (empathy explained almost half of the variance: $R^2 = .11$; the other half ($R^2 = .13$) was explained by gender, age, ethnicity, religiosity, conservative political orientation, and social class).

Moral inclination and empathy domains

Meta-analyses of the associations between empathic concern and the utilitarian and deontological parameters

We conducted two meta-analyses to examine the association between empathy (empathic concern) and the deontological and utilitarian parameters (two parameters that were derived from Conway and Gawronski's method, 2013). However, there were insufficient data on other moral inclination approaches or empathy domains to perform a meta-analysis.

[6]Participants had to decide between saving their job or their five colleagues' jobs.

[7]The participant in this dilemma has to choose between pushing someone out of a small ship to save six people (including the participant) or doing nothing and letting everyone die, including the participant.

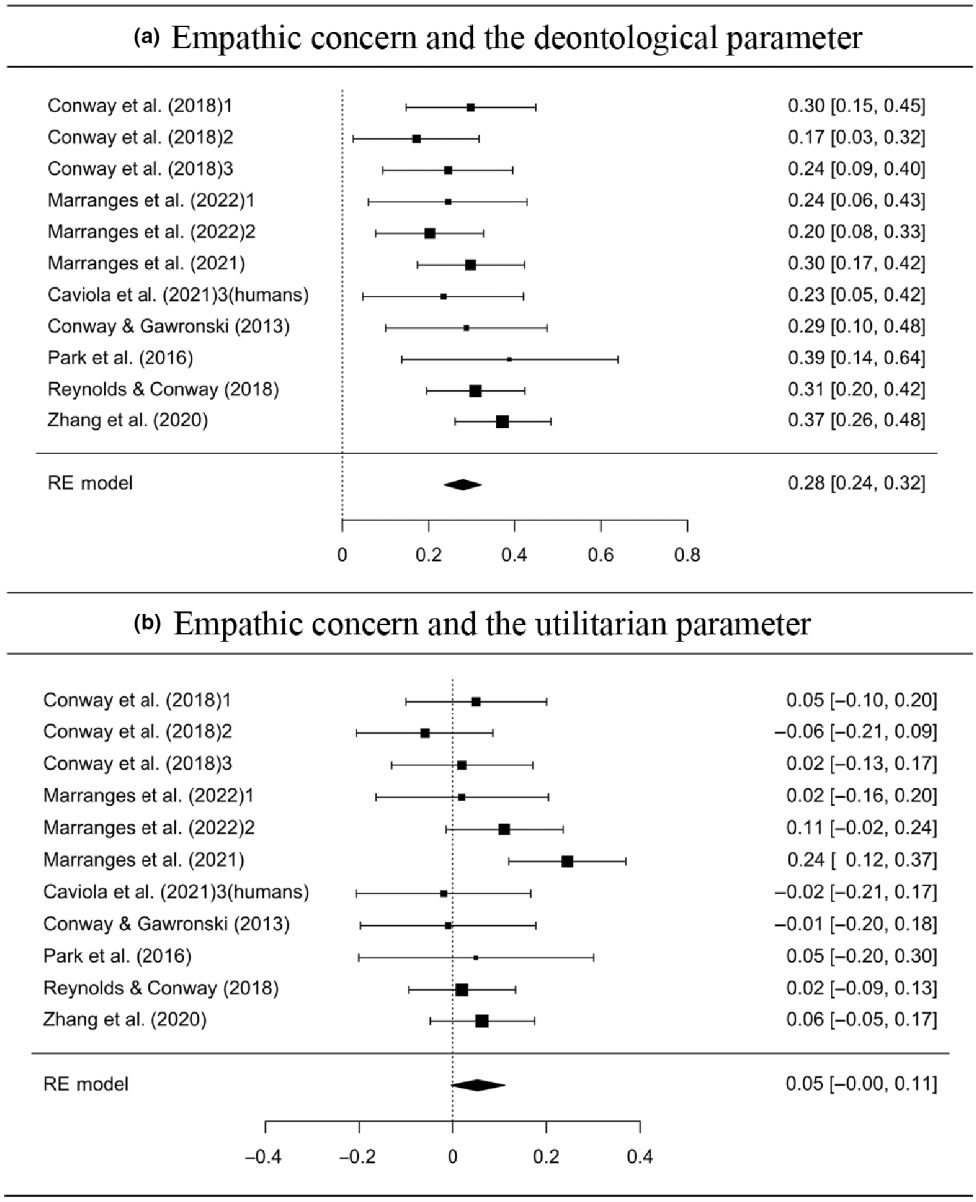


FIGURE 4 Forest plots of the meta-analytic correlations between empathy domains and moral inclination. (a) Empathic concern and the deontological parameter. (b) Empathic concern and the utilitarian parameter. Error bars indicate 95% CIs.

The first meta-analysis aimed to investigate the correlation between EC and the deontological parameter (DP). Using a random effect model ($k = 11$), we found a significant overall estimate ($E = 0.279$; $SE = 0.022$; $p < .001$; $CI^- = .236$; $CI^+ = .323$; see Figure 4a displaying the Forest plot), indicating a significant positive association between EC and DP. The Q-Test showed no significant amount of heterogeneity in the true outcomes, $Q_{(10)} = 7.87$; $p = .641$; $\tau^2 = 0$; $I^2 = 0\%$. None of the studies can be considered overly influential according to Cook's distances, and neither Beff and Mazumdar's rank correlation nor Egger's regression indicated any funnel plot asymmetry ($p = .542$; $p = .670$, respectively; see Figure S5a–c). Note that the fail-safe number was equal to 585 ($p < .001$), indicating no publication bias.

We conducted a second meta-analysis to examine the correlation between EC and the utilitarian parameter (UP). The random effects model ($k = 11$) displayed an insignificant overall estimate ($E = 0.054$; $SE = 0.026$; $p = .066$; $CI^- = -.004$; $CI^+ = .111$), indicating no association between EC and UP. Figure 4b displays the Forest plot illustrating this finding. The Q-Test showed no significant heterogeneity in the true outcomes, $Q_{(10)} = 13.8$; $p = .182$; $\tau^2 = 0.003$; $I^2 = 33.7\%$. However, one study (Maranges et al., 2021) was considered overly influential based on Cook's distances and had a value larger than |2.84| after examining the studentized residuals (see Figure S6a), suggesting an outlying study. Despite this, removing the outlier did not change the significance of the model ($E = 0.032$; $SE = 0.016$; $p = .078$), so we retained the outlying study in the random effects model. Neither Beff and Mazumdar's rank correlation nor Egger's regression indicated any funnel plot asymmetry in the $k = 11$ model ($p = .359$; $p = .276$, respectively). The funnel and Q–Q plots are displayed in Figure S6b,c. The fail-safe number was 10 ($p = .013$), indicating caution in interpreting the results of this meta-analysis.

A systematic review of the associations between empathy domains and moral inclination

The reviewed studies[5,10,12,13,16,25,27,29,31,34] examining the association between moral inclination and empathy employed three main methodologies: the Oxford Utilitarianism Scale[10,25], Conway and Gawronski's methodology[5,12,13,25,27,29,31,34], and intuitive utilitarian moral reasoning (a material leading more frequently to the utilitarian response)[16].

When using Conway and Gawronski's methodology (2013): (1) some authors[25] did not find significant associations between EC and the Deontological or Utilitarian Parameters (DP, UP; see Table 1). (2) Other authors[12] found small correlations between EC and the DP ($r = .20$; $r = .24$) and no association with the UP in their two studies. However, these findings were inconsistent, as multiple regression analyses revealed no significant predictive effects of EC on both UP and DP, while path analyses showed that EC significantly predicted both UP and DP (see Table 1). (3) In contrast, another reviewed study[13] reported a significant correlation between EC and DP ($r = .29$) and a small correlation between EC and UP ($r = .24$; see Table 1). (4) Other studies[27,29,31,34] showed moderate correlations between EC and the DP ($r = .28$[27]; $r = .37$[29] $r = .30$[31]; $r = .36$[34]; see Table 1), which were confirmed in multiple regression or mediation analyses. Additionally, an EC manipulation increased the deontological inclination[27]. Regarding other empathy domains, one studies[27] showed a moderate correlation between PT and the DP ($r = .32$), while another study[29] found no significant association between PT and the DP. However, neither EC nor PT was significantly associated with the UP[27,29], and the EC manipulation did not affect utilitarian inclination[27].

Studies using the Oxford Utilitarianism Scale (Kahane et al., 2015) found significant correlations between EC and the Oxford Utilitarianism Scale domains. Table 1 displays moderate correlations between EC and impartial concern for the greater good ($r = .33$[10]; Mean $r = .38$[25]) and small correlations between EC and permissive attitudes towards instrumental harm ($r = -.16$[10]; Mean $r = -.20$[25]).

Finally, authors[16] found a significant and positive predictive effect of EC on intuitive utilitarian moral reasoning among Korsakoff's syndrome patients (see Table 1). However, the study did not find any significant predictive effects of PT on intuitive utilitarian and deontological moral reasoning, and there were no significant predictive effects of EC on deontological moral reasoning.

Empathy for animals and utilitarian judgements, utilitarian choices, and moral inclination

One study[3] investigated the associations between empathy for animals and consequentialist or deontology preferences, as well as consequentialist decisions in a real-life mouse dilemma. Another study[25] examined the associations between EC and the permissibility of utilitarian actions in trolley-type scenarios

involving either animals or humans. In brief, only empathy for mice was significantly correlated with consequentialism preferences[3] ($r = -.14$, see Table 1). However, empathy for animals was not correlated with consequentialism or deontology preferences, and none of the empathy measures significantly predicted consequentialism decisions in the real-life trolley-type dilemma[3]. In the other reviewed study[25], no significant associations were found between EC and utilitarian or deontological inclinations when using animal trolley-type dilemmas (see Study 3 in Table 1).

DISCUSSION

The present systematic review and meta-analysis retained 34 papers investigating the relationship between empathy and moral judgements (judging the appropriateness of the utilitarian action), moral decision-making (utilitarian choices), or moral inclination (deontological or utilitarian preferences). These reviewed papers contain studies that were conducted all around the world, from America (north and south) to Europe, including some Asian countries, showing a large diversity of samples in terms of culture. Moreover, the retained papers were published in a ten-year period, revealing that these investigations are pretty recent.

Most of the included had large sample sizes and performed, on average, their investigations on young Caucasian adults (≈30 years old). Furthermore, the majority of the reviewed papers used the IRI (Davis, 1983) to measure empathy (≈80%), but other studies also used the BES (Jolliffe & Farrington, 2006), the affective and cognitive measure of empathy (Vachon & Lynam, 2016), or empathy-related adjectives. Concerning trolley materials, all authors used a large variety of trolley-type scenarios to measure utilitarian or deontological choices, judgements, or inclinations, including classical trolley problems, everyday trolley-like problems, utilitarian choices in an allocation money task (see Côté et al., 2013), and even real-life dilemmas (see Bostyn et al., 2018). Unfortunately, the first observation of the present work is that we are severely lacking in real-life trolley-type dilemmas presenting ethical conditions to be tested in the general or clinical populations.

Nevertheless, we believe this meta-analysis and systematic review will help clarify the associations between empathy and moral judgements, moral decision-making, or moral inclination and provide solid guidelines for future studies.

To what extent is empathy necessary in moral judgements, moral decision-making, and moral inclination?

Moral judgements and empathy domains

All in all, the meta-analysis and the systematic review convergently showed that affective empathy (from all empathy measurements) is negatively associated with utilitarian judgements (i.e., judging the appropriateness of the utilitarian action). The overall estimate from the meta-analysis showed a small association between these factors, but the systematic review displayed heightened associations in personal moral dilemmas (where the effect sizes can be categorized as moderate). Therefore, the more affective empathy, the less appropriate the utilitarian action. However, one cluster analysis showed surprising results by displaying that people who judged the utilitarian action as appropriate in a set of moral dilemmas had more EC than those who judged the utilitarian action as wrong (Rosas et al., 2019).

The other meta-analysis showed a small negative association between utilitarian judgements and cognitive empathy. Mostly, studies reported insignificant associations between cognitive empathy and utilitarian judgements, but three records[9,27,33] found significant correlations (two of them[27,33] reported especially high correlations compared to the others). However, multiple linear regressions extinguished these significant relationships. Hence, we can conclude that cognitive empathy has a limited and trivial role in how people judge utilitarian actions as appropriate (Figure 5 shows a summary of these results).

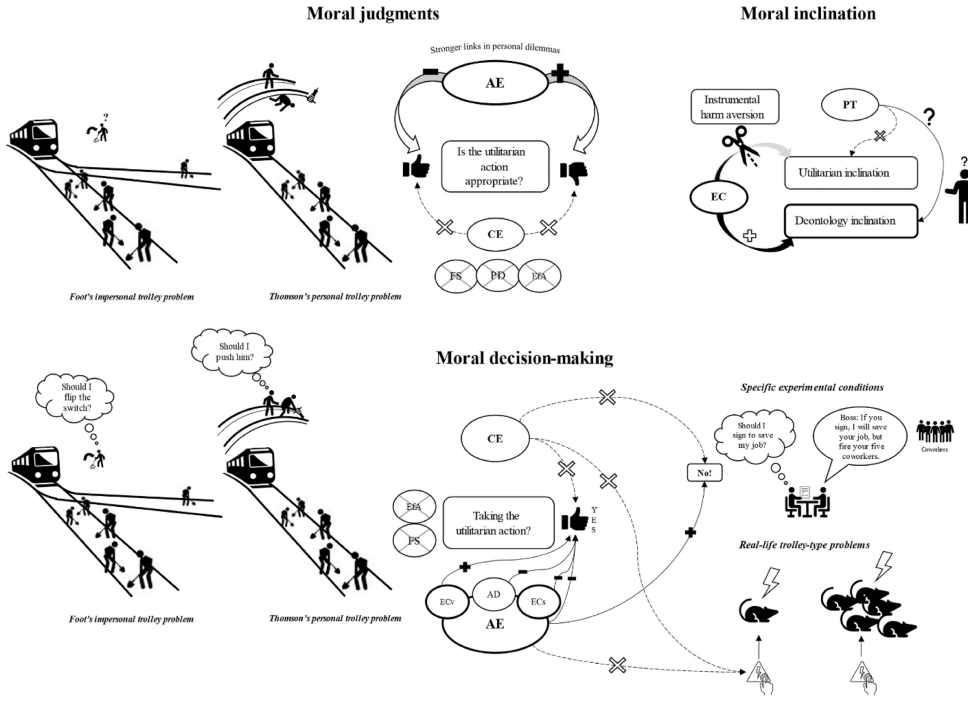


FIGURE 5 Summary illustration. *Note*: This figure depicts a summary of most findings of the reviewed records. The dotted lines represent limited associations. Abbreviations: AD, Affective Dissonance; AE, Affective Empathy; CE, Cognitive Empathy; EC, Empathic Concern; ECs, Empathic Concern for the Saved; ECv, Empathic Concern for the Victim; EfA, Empathy for Animals; FS, Fantasy; Instrumental harm aversion, Oxford Utilitarian Scale, Permissive attitude towards instrumental harm; PD, Personal Distress; PT, Perspective-Taking.

Finally, concerning other empathy subdomains, a clear conclusion can be drawn based on the reviewed papers: FS and PD are not significantly related to utilitarian judgements. The rare significant correlations or predictive effects were minimal in terms of effect sizes or sample sizes, suggesting anecdotal results. Figure 5 shows a summary of these results.

Moral decision-making and empathy domains

The first meta-analysis revealed a small and negative association between affective empathy and utilitarian choices (the more affective empathy, the less utilitarian choices). When using the IRI, most reviewed studies showed that EC was negatively associated with utilitarian choices, and some authors only found this association in personal moral dilemmas. In the same direction, other empathy approaches (BES) showed a significant predictive effect, albeit small, of AE on utilitarian choices. However, this effect was found under specific experimental conditions[14,28] or varied according to gender[14]. Note that another study found no significant links between the BES domains and utilitarian choices[32]. Interestingly, EC most strongly classified utilitarian decision-makers apart using a discriminant analysis. One last study[4] showed that empathy (mainly referring to EC) negatively and importantly predicted utilitarian allocations when using empathy adjectives and a utilitarian monetary task. Overall, the reported effect sizes of the association between EC and utilitarian choices were mainly small and moderate in a few cases. A summary of these results is illustrated in Figure 5.

The other meta-analysis showed that cognitive empathy was not significantly associated with utilitarian choices. Most studies using the IRI did not display any significant association or significant

differences between utilitarian and deontological responders in their PT scores, and only sparse results displayed some significant effect, suggesting anecdotal findings. The most convincing findings emerged from one study using PT induction for the losing member and another using a real-life trolley-type dilemma: both showed no significant impact or association between PT (and other IRI domains) and utilitarian choices. However, when using other empathy measurements (i.e., empathy adjectives), some reported some negative predictive effects of cognitive empathy on utilitarian choices in two everyday trolley-like problems (and a negative predictive effect of affective empathy on utilitarian choices in ostracism everyday trolley-like problem[33]).

Regarding other empathy domains or approaches, despite some anecdotal findings (i.e., two studies[22,30] reported significant negative associations between PD and utilitarian choices in personal moral dilemmas, and one study[20] reported a significant and positive relation between FS and utilitarian choices), most studies found insignificant associations between PD or FS and utilitarian choices. When authors used some empathy adjectives to measure empathy[4,23,33], one author found small but significant predictive effects of AE or CE on utilitarian choices[33]. In an everyday trolley-like problem involving ostracism, both AE and CE negatively predicted utilitarian choices, whereas only CE negatively predicted utilitarian choices in a job termination everyday trolley-like problem. Finally, the authors subdivided empathy into other-focused (i.e., empathic concern) and self-focused indices (i.e., personal distress) and found that other and self-focused empathy negatively predicted utilitarian choices in the footbridge dilemma and positively predicted utilitarian choices in the raft-boat dilemma (see Table 1). Another study using one atypical empathy approach (i.e., the affective and cognitive measure of empathy; Vachon & Lynam, 2016) showed that affective dissonance (a new empathy domain that has no equivalent on other measurement scales which refers to antagonizing emotional reactions and responses, like feeling enjoyment or pleasure from others' misfortune or pain) significantly predicted utilitarian choices in both personal and impersonal scenarios (in a negative way). The reason why this domain of empathy (assuming that it is indeed a domain of empathy) is a negative predictor of utilitarian choices remains unclear.

Finally, although the magnitude of these predictors is unknown, some authors[23] have shown that empathic concern for the victim negatively predicted utilitarian choices in the footbridge and raftboat dilemmas (i.e., higher EC for the victim predicted less utilitarian choices), and empathic concern for the saved positively predicted utilitarian choices in both of these dilemmas (i.e., higher EC for the saved predicted more utilitarian choices).

Moral inclination and empathy domains

The meta-analyses indicated a significant positive association between EC (all studies focusing on moral inclination used the IRI) and deontological inclinations, but no significant association was found between EC and utilitarian inclinations. Most of the studies reviewed reported that EC was significantly correlated (with small to moderate correlations) with deontological inclinations but not with utilitarian inclinations, meaning that as EC increased, deontological inclinations also increased. Importantly, one study[27] showed that manipulating EC selectively increased deontological inclinations but not utilitarian inclinations. However, one study[13] reported a small significant link between EC and UP and another[25,study3] found no significant links between EC and deontological or utilitarian inclinations (although there was a correlation of .23 between EC and deontological inclinations). Finally, one study[16] using a utilitarian moral reasoning approach showed a positive predictive effect of EC on utilitarian moral reasoning but not on deontological moral reasoning.

Regarding the links between PT and utilitarian or deontological inclinations, two studies[27,29] have produced contrasting results: one study found a moderate positive correlation between PT and deontological inclinations[27], while the other found no significant association[29]. However, both studies consistently reported no association between PT and utilitarian inclinations (which is in line with the results found by other authors who have used a utilitarian moral reasoning approach[16]).

Finally, when using the Oxford Utilitarianism Scale, two studies[10,25] reported approximately the same patterns of correlations: EC was moderately and positively associated with impartial concern for the greater good, and EC presented a (small/moderate) negative association with permissive attitudes towards instrumental harm. The authors explained these results by stating: "although people who feel greater empathic concern care more about impartially maximizing welfare, they are also less likely to endorse instrumental harm to achieve those ends."[10, p.152]. This explanation shows how complex the association between empathy and morality is. Figure 5 illustrates a summary of these findings.

Note that none of the mentioned studies explored the association between the other empathy subdomains and moral inclination.

Empathy for animals and trolley problems and variants

Two studies reviewed in this analysis[3,25] examined potential links between empathy for animals and moral judgement, moral decision-making, or moral inclination. However, only one study reported a negative correlation between empathy for mice and consequentialist preferences, which was merely anecdotal. Based on the small effect size and lack of significant relationships found in these studies, we can conclude that empathy for animals is unrelated to moral judgement, moral decision-making, or moral inclination, even when using real-life trolley problems.

GENERAL DISCUSSION

In this systematic review and meta-analysis, our goal was to clarify the role of empathy domains in moral judgement, moral decision-making, and moral inclination when using trolley problems and variants. As outlined in the introduction, some experts in empathy, such as Jean Decety and Frans de Waal, hold opposing viewpoints. Frans de Waal believes empathy is a crucial component of morality (as also argued by Rifkin, 2009). In contrast, Jean Decety contends that while empathy may be essential, it is not necessarily the most powerful predictor of moral judgement and decision-making. Although trolley problems and variants only provide a partial approach to morality, they remain one of the most significant materials in research that could help clarify these differing positions.

The reviewed papers provide evidence that affective empathy is negatively associated with utilitarian judgements and choices, with a small to moderate effect size. Most reviewed studies found that the higher levels of EC, the less appropriate the utilitarian action and the fewer utilitarian choices. This suggests that the salient state of vulnerability or need of the dilemma's protagonists elicits specifically empathic concern feelings, as proposed by Baston (2012). Furthermore, regarding moral inclination, we found that EC was positively related to the deontological parameter when using Conway and Gawronski's methodology but not to the utilitarian parameter. Interestingly, using the Oxford Utilitarianism Scale (Kahane et al., 2015) revealed a new understanding of the complex association between EC and utilitarianism. Some authors[10,25] demonstrated that individuals with higher EC levels display a more impartial utilitarian inclination, yet are less capable of performing instrumental harm to pursue this impartial inclination. In terms of other empathy domains, such as cognitive empathy, PT, FS, or PD, most reviewed studies showed no significant (or limited) associations with utilitarian judgements and choices. Although some sparse results displayed significant associations, these might be statistical artefacts due to small samples or effect sizes. Importantly, PT was unrelated to utilitarian choices, especially when using a PT induction for the losing member or a real-life trolley-type dilemma. While it is clear that PT is not related to the utilitarian parameter (moral inclination), it remains unclear whether PT has a significant relation with the deontological parameter, as studies reported divergent results.

Based on the meta-analyses and the papers that were systematically reviewed, it was found that empathy, particularly affective empathy, is significantly associated with various facets of morality, such

as moral judgements, moral decision-making, and moral inclination. However, the overall effect sizes indicate that this relation is small (although it was relatively stronger at the level of moral inclination). Therefore, empathy, including affective empathy, is not such a fundamental pillar of morality as proposed by Frans de Waal. Instead, our findings support Jean Decety's perspective that empathy is likely essential in morality but not the most relevant predictor, as affective and cognitive empathy have differing impacts on morality.

Furthermore, the present systematic review and meta-analysis revealed surprising results that do not uniformly support the Dual-Process Model (Greene et al., 2004). Here, it is worth noting that we are drawing a parallel between the domains of empathy and the facets of the Dual-Process Model (DPM) proposed by Greene et al. (2004). Yu and Chou (2018) have described empathy as a dual-route model, with one route being fast and involving the mirror neuron system (affective empathy), while the other route (cognitive empathy) is slower and engages brain regions such as the precuneus, temporoparietal junction, and the dorsal, middle, and ventral medial prefrontal cortex. However, while the dual-route model of empathy bears some resemblance to the DPM, it is primarily aimed at explaining prosocial behaviours, not moral judgements (the DPM was initially proposed as a model for moral judgements but later extended to other moral facets like moral decision-making). In fact, the DPM suggests that moral judgements involve both affective and cognitive processes (Greene & Haidt, 2002). Our findings indicate that cognitive empathy (i.e., a cognitive process) has limited associations with moral judgements and decision-making (associations that disappeared when using more complex statistical analyses), and other findings seem to support this conclusion. For instance, studies on alexithymia (i.e., the difficulty in identifying and finding words to describe emotions; Nemiah, 1976) and moral judgements displayed no significant association between the externally oriented way of thinking (an alexithymia domain described as a pragmatical way of thinking that is expected to shape cognitive processes, also called *pensée opératoire*) and judging the utilitarian action as appropriate (Takamatsu & Takai, 2019). Hence, it is suggested that the processes underlying moral judgements, moral decision-making, and moral inclinations may be more complex than the dichotomous approach proposed by the DPM. Such discrepancies between theory and empirical evidence highlight the need to redefine the concepts of "affective and cognitive processes" that were assumed to play a role in morality (e.g., What are the specific characteristics associated with affective and cognitive processes? Do we have clear enough criteria that allow us to define and distinguish these two processes?). This could be achieved through the development of new models or the refinement of existing models.

In conclusion, we present the first systematic review and meta-analysis analysing the associations between empathy and moral judgements, moral decision-making, and moral inclination in trolley problems and their variants. The reviewed studies employed interesting methodologies, such as examining empathic concern towards the victim or the saved (Takamatsu, 2018), exploring different protagonist perspectives (Nasello et al., 2021, 2023), manipulating empathic concern or perspective-taking (Conway & Gawronski, 2013; Côté et al., 2013), investigating utilitarianism for humans or animals (Bostyn et al., 2018; Caviola et al., 2021), and developing real-life trolley-type dilemmas (Bostyn et al., 2018). Therefore, future research using these methodologies, especially the development of ethical versions of real-life trolley-type problems involving humans, would likely benefit the field and yield more realistic results regarding moral judgements, moral decision-making, and moral inclination. Finally, we believe this systematic review and meta-analysis will offer a clear nomenclature for the different facets of morality (i.e., moral judgements, moral decision-making, and moral inclination) to avoid any confusion between these distinct processes.

AUTHOR CONTRIBUTIONS

Julian A. Nasello: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; software; validation; visualization; writing – original draft; writing – review and editing. **Jean-Marc Triffaux:** Conceptualization; project administration; resources; supervision; validation; visualization; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENTS
None.

CONFLICT OF INTEREST STATEMENT
The authors declare no competing interests.

DATA AVAILABILITY STATEMENT
The online data file is available on the following DOI link: https://osf.io/u382h/.

ORCID
Julian A. Nasello https://orcid.org/0000-0003-0188-1025
Jean-Marc Triffaux https://orcid.org/0000-0002-0058-6190

REFERENCES
Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Baez, S., Flichtentrei, D., Prats, M., Mastandueno, R., García, A. M., Cetkovich, M., & Ibáñez, A. (2017). Men, women… who cares? A population-based study on sex differences and gender roles in empathy and moral cognition. *PLoS One*, *12*(6), e0179336. https://doi.org/10.1371/journal.pone.0179336

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*, 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Baheti, A. D., & Nayak, P. (2022). Covid-19 in India: Oxygen shortages and a real world trolley problem. *BMJ*, *376*, o369. https://doi.org/10.1136/bmj.o369

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*(1), 154–161. https://doi.org/10.1016/j.cognition.2011.05.010

Baston, C. D. (2012). The empathy-altruism hypothesis: Issues and implication. In J. Decety (Ed.), *Empathy: From bench to bedside*. MIT Press.

Behnke, A., Strobel, A., & Armbruster, D. (2020). When the killing has been done: Exploring associations of personality with third-party judgment and punishment of homicides in moral dilemma scenarios. *PLoS One*, *15*(6), e0235253. https://doi.org/10.1371/journal.pone.0235253

Benlahcene, A., Zainuddin, R. B., Syakiran, N., & Ismail, A. B. (2018). A narrative review of ethics theories: Teleological & deontological ethics. *Journal of Humanities and Social Science (IOSR-JHSS)*, *23*(1), 31–32. https://doi.org/10.9790/0837-2307063138

Bentham, J. (1789). *Introduction to the principles of morals and legislation*. Payne.

Bermond, B., Oosterveld, P., & Vorst, H. C. M. (1994). *Bermond-Vorst Alexithymia Questionnaire; construction, reliability, validity and uni-dimensionality*. Internal Report. University of Amsterdam: Faculty of Psychology, Department of Psychological Methods.

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, *29*(7), 1084–1093. https://doi.org/10.1177/0956797617752640

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413–425. https://doi.org/10.2307/1128984

Burghart, M., & Mier, D. (2022). No feelings for me, no feelings for you: A meta-analysis on alexithymia and empathy in psychopathy. *Personality and Individual Differences*, *194*, 111658. https://doi.org/10.1016/j.paid.2022.111658

Caviola, L., Kahane, G., Everett, J. A., Teperman, E., Savulescu, J., & Faber, N. S. (2021). Utilitarianism for animals, Kantianism for people? Harming animals and humans for the greater good. *Journal of Experimental Psychology: General*, *150*(5), 1008–1039. https://psycnet.apa.org/doi/10.1037/xge0000988

Cecchetto, C., Korb, S., Rumiati, R. I., & Aiello, M. (2018). Emotional reactions in moral decision-making are influenced by empathy and alexithymia. *Social Neuroscience*, *13*(2), 226–240. https://doi.org/10.1080/17470919.2017.1288656

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–235. https://psycnet.apa.org/doi/10.1037/a0031021

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition*, *179*, 241–265. https://doi.org/10.1016/j.cognition.2018.04.018

Côté, S., Piff, P. K., & Willer, R. (2013). For whom do the ends justify the means? Social class and utilitarian moral judgment. *Journal of Personality and Social Psychology*, *104*(3), 490–503. https://psycnet.apa.org/doi/10.1037/a0030931

Craig, S. C., McCarthy, A. F., & Gainous, J. (2022). Question wording and attitudinal ambivalence: COVID, the economy, and Americans' response to a real-life trolley problem. *Social Science Quarterly*, *103*(1), 18–30. https://doi.org/10.1111/ssqu.13102

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/2F1754073914558466

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalogue of Selected Documents in Psychology*, *10*, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://psycnet.apa.org/doi/10.1037/0022-3514.44.1.113

de Waal, F., Macedo, S. E., & Ober, J. E. (2006). *Primates and philosophers: How morality evolved*. Princeton University Press.

Decety, J. (2021). Why empathy is not a reliable source of information in moral decision making. *Current Directions in Psychological Science*, *30*(5), 425–430. https://doi.org/10.1177/2F09637214211031943

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science*, *9*(5), 525–537. https://doi.org/10.1177/1745691614545130

Decety, J., & Cowell, J. M. (2015). Empathy, justice, and moral behavior. *AJOB Neuroscience*, *6*(3), 3–14. https://doi.org/10.1080/21507740.2015.1047055

Dinić, B. M., Milosavljević, M., & Mandarić, J. (2021). Effects of Dark Tetrad traits on utilitarian moral judgement: The role of personal involvement and familiarity with the victim. *Asian Journal of Social Psychology*, *24*(1), 48–58. https://doi.org/10.1111/ajsp.12422

Eklund, J. H., & Meranius, M. S. (2021). Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling*, *104*(2), 300–307. https://doi.org/10.1016/j.pec.2020.08.022

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. In P. Foot (Ed.), (1977) *Virtues and vices and other essays in moral philosophy* (pp. 19–32). Basil Blackwell.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology*, *113*(3), 343–376. https://doi.org/10.1037/pspa0000086

Gleichgerrcht, E., Torralva, T., Rattazzi, A., Marenco, V., Roca, M., & Manes, F. (2013). Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. *Social Cognitive and Affective Neuroscience*, *8*(7), 780–788. https://doi.org/10.1093/scan/nss067

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS One*, *8*(4), e60418. https://doi.org/10.1371/journal.pone.0060418

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*(12), 517–523. https://doi.org/10.1016/S1364-6613(02)02011-9

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J. (2003). The Moral Emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*. Oxford University Press.

Hannikainen, I. R., Machery, E., & Cushman, F. A. (2018). Is utilitarian sacrifice becoming more morally permissible? *Cognition*, *170*, 95–101. https://doi.org/10.1016/j.cognition.2017.09.013

Hepper, E. G., Hart, C. M., & Sedikides, C. (2014). Moving Narcissus: Can narcissists be empathic? *Personality and Social Psychology Bulletin*, *40*(9), 1079–1091. https://doi.org/10.1177/0146167214

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*, 307–316. https://psycnet.apa.org/doi/10.1037/h0027580

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas? *PLoS One*, *11*(7), e0154780. https://doi.org/10.1371/journal.pone.0154780

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence*, *29*, 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience*, *10*(5), 551–560. https://doi.org/10.1080/17470919.2015.1023400

Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. https://doi.org/10.1037/rev0000093

Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, *134*, 193–209. https://doi.org/10.1016/j.cognition.2014.10.005

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402. https://doi.org/10.1093/scan/nsr005

Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten*. Johann Friedrich Hartknoch.

Kogut, T., & Ritov, I. (2005). The "identified victim" effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*(3), 157–167. https://doi.org/10.1002/bdm.492

Maranges, H. M., Chen, S. K., & Conway, P. (2022). Insecure and insensitive: Avoidant and anxious attachment predict less concern for others in sacrificial moral dilemmas. *Personality and Individual Differences*, *185*, 111274. https://doi.org/10.1016/j.paid.2021.111274

Maranges, H. M., Hasty, C. R., Maner, J. K., & Conway, P. (2021). The behavioral ecology of moral dilemmas: Childhood unpredictability, but not harshness, predicts less deontological and utilitarian responding. *Journal of Personality and Social Psychology*, *120*(6), 1696–1719. https://doi.org/10.1037/pspp0000368

May, J. (2018). The limits of emotion in moral. In K. Jones & F. Schroeter (Eds.), *The many moral rationalisms* (pp. 286–306). Oxford University Press.

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology*, *18*(4), 193–197. https://psycnet.apa.org/doi/10.1097/01.wnn.0000191292.17964.bb

Mill, J. S. (1861). Utilitarianism. In J. M. Robson (Ed.), *The collected works of John Stuart Mill. Essays on ethics, religion and society* (Vol. *X*, pp. 203–259). University of Toronto Press.

Mitchell, M. B., & Attipoe, G. M. (2020). Revisiting the "Trolley Problem" in COVID-19 pandemic. *Hektoen International: A Journal of Medical Humanities*. https://hekint.org/2020/10/15/revisiting-the-trolley-problem-in-the-covid-19-pandemic/

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*, 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549–557. https://doi.org/10.1111/j.1467-9280.2008.02122.x

Nakamura, H., Ito, Y., Honma, Y., Mori, T., & Kawaguchi, J. (2014). Cold-hearted or cool-headed: Physical coldness promotes utilitarian moral judgment. *Frontiers in Psychology*, *5*, 1086. https://doi.org/10.3389/fpsyg.2014.01086

Nasello, J. A., Dardenne, B., Blavier, A., & Triffaux, J. M. (2021). Does empathy predict decision-making in everyday trolley-like problems? *Current Psychology*, *1-14*, 2966–2979. https://doi.org/10.1007/s12144-021-01566-1

Nasello, J. A., Dardenne, B., Hansenne, M., Blavier, A., & Triffaux, J.-M. (2023). Moral decision-making in trolley problems and variants: how do participants' perspectives, borderline personality traits, and empathy predict choices? *The Journal of Psychology: Interdisciplinary and Applied*, 1–21. https://doi.org/10.1080/00223980.2023.2206604

Nemiah, J. C. (1976). Alexithymia: A view of the psychosomatic process. *Modern Trends in Psychosomatic Medicine*, *3*, 430–439.

Oudman, E., van Stigt Thans, S., Montoya, E. R., & Postma, A. (2021). Moral reasoning, moral decision-making, and empathy in Korsakoff's syndrome. *Journal of Neuropsychology*, *15*(3), 462–476. https://doi.org/10.1111/jnp.12233

Oveis, C., Horberg, E. J., & Keltner, D. (2010). Compassion, pride, and social intuitions of self-other similarity. *Journal of Personality and Social Psychology*, *98*(4), 618–630. https://psycnet.apa.org/doi/10.1037/a0017628

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMC Systematic Reviews*, *10*, 89. https://doi.org/10.1186/s13643-021-01626-4

Park, G., Kappes, A., Rho, Y., & Van Bavel, J. J. (2016). At the heart of morality lies neuro-visceral integration: Lower cardiac vagal tone predicts utilitarian moral judgment. *Social Cognitive and Affective Neuroscience*, *11*(10), 1588–1596. https://doi.org/10.1093/scan/nsw077

Patil, I., Larsen, E. M., Kichic, R., & Gleichgerrcht, E. (2021). Moral cognition about harm in anxiety disorders: The importance of experienced emotion. *Psychological Reports*, *124*(6), 2501–2523. https://doi.org/10.1177/0033294120964134

Patil, I., Melsbach, J., Hennig-Fast, K., & Silani, G. (2016). Divergent roles of autistic and alexithymic traits in utilitarian moral judgments in adults with autism. *Scientific Reports*, *6*(1), 1–15. https://doi.org/10.1038/srep23637

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology*, *5*, 501. https://doi.org/10.3389/fpsyg.2014.00501

Paul, E. S. (2000). Empathy with animals and with humans: Are they linked? *Anthrozoös*, *13*(4), 194–202. https://doi.org/10.2752/089279300786999699

Rifkin, J. (2009). *The empathic civilization: The race to global consciousness in a world in crisis*. Penguin Group.

Romero-Rivas, C., López-Benítez, R., & Rodríguez-Cuadrado, S. (2022). Would you sacrifice yourself to save five lives? Processing a foreign language increases the odds of self-sacrifice in moral dilemmas. *Psychological Reports*, *125*(1), 498–516. https://doi.org/10.1038/srep23637

Rosas, A., Viciana, H., Caviedes, E., & Arciniegas, A. (2019). Hot utilitarianism and cold deontology: Insights from a response patterns approach to sacrificial and real world dilemmas. *Social Neuroscience*, *14*(2), 125–135. https://doi.org/10.1080/17470919.2018.1464945

Rota, G., Palumbo, S., Lattanzi, N., Manfrinati, A., Sarlo, M., Lotto, L., Pietrini, P., Rumiati, R., & Pellegrini, S. (2016). Harm aversion explains utilitarian choices in moral decision-making in males but not in females. *Archives Italiennes de Biologie*, *154*, 50–58. https://doi.org/10.12871/000398292

Santamaría-García, H., Baez, S., García, A. M., Flichtentrei, D., Prats, M., Mastandueno, R., Sigman, M., Matallana, D., Cetkovich, M., & Ibáñez, A. (2017). Empathy for others' suffering and its mediators in mental health professionals. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-06775-y

Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don't do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology & Behavior*, *130*, 127–134. https://doi.org/10.1016/j.physbeh.2014.04.002

Shao, C. (2020). The COVID trolley dilemma. *The American Journal of Surgery*, *220*(3), 545–549. 10.1016%2Fj.amjsurg.2020.05.012

Sidgwick, H. (1874). *The methods of ethics*. Macmillan.

Smillie, L. D., Katic, M., & Laham, S. M. (2020). Personality and moral judgment: Curious consequentialists and polite deontologists. *Journal of Personality*, *89*(3), 549–564. https://doi.org/10.1111/jopy.12598

Takamatsu, R. (2018). Turning off the empathy switch: Lower empathic concern for the victim leads to utilitarian choices of action. *PLoS One*, *13*(9), e0203826. https://doi.org/10.1371/journal.pone.0203826

Takamatsu, R. (2019). Personality correlates and utilitarian judgments in the everyday context: Psychopathic traits and differential effects of empathy, social dominance orientation, and dehumanization beliefs. *Personality and Individual Differences*, *146*, 1–8. https://doi.org/10.1016/j.paid.2019.03.029

Takamatsu, R., & Takai, J. (2019). With or without empathy: Primary psychopathy and difficulty in identifying feelings predict utilitarian judgment in sacrificial dilemmas. *Ethics & Behavior*, *29*(1), 71–85. https://doi.org/10.1080/10508422.2017.1367684

Underwood, L. G., & Teresi, J. A. (2002). The daily spiritual experience scale: Development, theoretical description, reliability, exploratory factor analysis, and preliminary construct validity using health-related data. *Annals of Behavioral Medicine*, *24*(1), 22–33. https://doi.org/10.1207/S15324796ABM2401_04

Vachon, D. D., & Lynam, D. R. (2016). Fixing the problem with empathy: Development and validation of the affective and cognitive measure of empathy. *Assessment*, *23*(2), 135–149. https://doi.org/10.1177/10731911145

Vyas, K., Jameel, L., Bellesi, G., Crawford, S., & Channon, S. (2017). Derailing the trolley: Everyday utilitarian judgments in groups high versus low in psychopathic traits or autistic traits. *Psychiatry Research*, *250*, 84–91. https://doi.org/10.1016/j.psychres.2017.01.054

Yu, C. L., & Chou, T. L. (2018). A dual route model of empathy: A neurobiological prospective. *Frontiers in Psychology*, *9*, 2212. https://doi.org/10.3389/fpsyg.2018.02212

Zhang, X., Wu, Z., Li, S., Lai, J., Han, M., Chen, X., Liu, C., & Ding, D. (2020). Why people with high alexithymia make more utilitarian judgments: The role of empathic concern and deontological inclinations. *Experimental Psychology*, *67*(1), 23–30. https://doi.org/10.1027/1618-3169/a000474

Zhou, X., Wildschut, T., Sedikides, C., Shi, K., & Feng, C. (2012). Nostalgia: The gift that keeps on giving. *Journal of Consumer Research*, *39*(1), 39–50. https://doi.org/10.1086/662199

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nasello, J. A., & Triffaux, J.-M. (2023). The role of empathy in trolley problems and variants: A systematic review and meta-analysis. *British Journal of Social Psychology*, *00*, 1–29. https://doi.org/10.1111/bjso.12654

Synthèse de la méta-analyse et revue systématique :

Les résultats de cette méta-analyse dresse une conclusion claire : c'est principalement l'empathie affective qui est significativement associée aux trois champs de la morale dans les dilemmes de type trolley.

Au niveau des jugements et choix moraux, les associations (faibles) apparaissent négatives, ce qui signifie que plus un individu va présenter des scores d'empathie affective élevés, moins il va avoir tendance à réaliser un jugement ou un choix utilitaire. Concernant l'inclinaison morale, on voit très clairement qu'il y a une association positive modérée entre la préoccupation empathique (une sous-dimension d'empathie affective) et le paramètre déontologique. Dès lors, un individu ayant des scores de préoccupation empathique élevés va davantage « s'orienter » vers une tendance déontologique (où l'action ne va pas violer les interdits moraux). Toutefois, Kahane et collaborateurs (2012 ; 2015 ; 2018) ont apporté une nuance capitale à ce résultat. Ils ont montré que les individus ayant des scores de préoccupation empathique élevés ont une tendance « utilitariste » (i.e., ils vont viser à maximiser de manière impartiale le bien-être commun), tout en présentant une aversion marquée à faire du mal intentionnellement aux autres (*instrumental harm aversion*). De ce fait, il semblerait que c'est l'aversion marquée à faire du mal intentionnellement aux autres qui dirige les individus avec une préoccupation empathique élevée à présenter une inclinaison déontologique, malgré le fait qu'ils adhèrent à l'idée qu'il faille maximiser de façon impartiale le bien-être commun (position utilitariste). En d'autres termes, cette aversion à faire du mal intentionnellement à autrui semble refreiner l'adhérence à la tendance utilitariste, ce qui produit une tendance déontologique.

Concernant l'empathie cognitive, les associations significatives apparaissent majoritairement triviales. Effectivement, les associations significatives ayant émergé disparaissent assez rapidement lorsque l'on utilise des méthodologies statistiques qui prennent en compte d'autres variables (typiquement, via des analyses de régressions linéaires multiples).

En résumé, l'empathie est-elle un facteur clé des jugements et choix moraux ou de l'inclinaison morale ? Les résultats de notre méta-analyse ont démontré que, dans le cadre des dilemmes de type trolley, l'empathie (surtout la dimension affective) peut être, sous

certains aspects, une variable significativement associée aux jugements et choix moraux et à l'inclinaison morale. Toutefois, elle n'apparaît pas comme la variable explicative la plus pertinente et prédominante pour expliquer les choix et jugements moraux ou l'inclinaison morale (comme soutenu par Decety, 2021 ; Decety & Cowell, 2014 ; 2015). Autrement dit, l'empathie est loin d'incarner le rôle qu'on lui confère dans la morale, à savoir celui d'être l'un des deux piliers principaux de la morale avec la réciprocité (Cf., de Waal et al., 2006).

Le DPM (Greene et al., 2001 ; 2004), une théorie qui a du plomb dans l'aile ?

L'empathie, avec ses dimensions affectives et cognitives, est un facteur de choix pour tester les prédictions que l'on peut inférer à partir du DPM. En effet, l'empathie regroupe les deux facettes centrales du DPM et, avec le modèle de l'empathie développé par Yu et Chou (2018), nous pouvons réaliser un parallèle entre ces deux modèles. D'après Yu et Chou (2018), l'empathie est décrite selon deux voies. Une première voie rapide qui implique le système des neurones miroirs ; cette voie est propre à la dimension affective de l'empathie. L'autre voie est plutôt lente et implique des régions cérébrales comme le précunéus, la jonction temporopariétale et le cortex préfrontal (dorsal, ventral et médian) ; cette voie est propre à la dimension cognitive de l'empathie. La différence majeure entre les deux modèles (empathie et DPM) est que l'un vise à expliquer les comportements prosociaux (l'empathie) et l'autre les jugements moraux (le DPM).

À travers nos trois études principales et la méta-analyse et revue systématique que nous avons présenté, nous avons pu montrer que le pouvoir explicatif et prédictif du DPM est relativement limité. De plus, certains résultats restent inexpliqués lorsque l'on se base sur les éléments théoriques de ce modèle. Par exemple, l'absence d'effet significatif (consistant) de l'empathie cognitive pour expliquer les jugements et choix moraux. Ou encore, lorsque d'autres paramètres pouvant être associés aux « processus cognitifs » (e.g., la pensée opératoire de l'alexithymie et l'intellect) ne sont pas (ou que faiblement) associés aux jugements moraux (Takamatsu & Takai, 2019 ; Smillie et al., 2020). En ce qui concerne les processus émotionnels du DPM, les données empiriques soutiennent dans l'ensemble que le modèle surévalue l'importance donnée aux émotions pour expliquer les jugements moraux (e.g., Horne & Powell, 2016). Ces auteurs ont montré que les dilemmes suscitaient, effectivement, des réponses émotionnelles fortes. Toutefois, ces dernières n'apparaissaient que faiblement associées aux jugements moraux. Ceci rejoint les effets obtenus avec l'empathie affective où l'on voit une association significative avec les jugements et choix moraux et l'inclinaison morale, mais cette association reste faible dans la plupart des cas (Nasello & Triffaux, 2023b). Dès lors, comme l'ont souligné nos travaux (Nasello et al., 2021a ; Nasello et al., 2023a ; Nasello & Triffaux, 2023b ; Nasello & Triffaux, *In press*) et ceux d'autres auteurs (Duke & Bègue, 2015 ; Horne & Powell, 2020 ;

Smillie et al., 2021 ; Decety, 2021 ; Oudman et al., 2021), les différents éléments de preuves soutiennent l'idée que ce modèle est probablement incomplet ou incorrect. Pourtant, à ce jour, le DPM demeure le modèle de référence. Effectivement, selon Google Scholar, les quatre articles clés de Greene et collaborateurs (Greene et al., 2001 ; 2004 ; 2007 ; Greene, 2009) qui ont contribué à la mise au point du DPM comptabilisent 10.449 citations, ce qui surpasse de très loin les alternatives à ce modèle.

Toutefois, en juillet 2023, Joshua Greene publie un article théorique soutenant que le DPM, bien que correct dans ses fondations, devrait être révisé. En effet, cet auteur phare maintient, en se basant sur des données issues de patients cérébrolésés, qu'il y a bien des aspects cognitifs et émotionnels qui vont produire certains types de réponses (i.e., davantage de jugements déontologiques ou utilitaires, comme expliqué au premier chapitre), mais il ne serait plus véritablement question de processus rapides et automatiques (émotionnels) versus des processus lents et réflexifs (cognitifs). Ce changement survient à la suite de données montrant que les réponses utilitaires ne sont pas lentes (Bago & De Neys, 2019 ; Baron & Gürçay, 2017 ; Cova et al., 2021 ; Gürçay & Baron, 2017 ; Koop, 2013 ; Rosas et al., 2019 ; Tinghög et al., 2016) ; contrairement à ce que d'autres auteurs (dont Greene) avaient montré (Greene et al., 2008 ; Suter & Hertwig, 2011 ; Trémolière et al., 2012). Les auteurs ayant démontré que les réponses utilitaires ne sont pas lentes soutiennent que les deux types de réponses (utilitaires ou déontologiques) sont intuitives (i.e., elles peuvent survenir rapidement chez le décideur) et qu'elles ne sont pas sous-tendues par des processus en opposition.

En résumé, les explications et recommandations demeurent assez maigres et aucune modification tangible n'a été réalisée à ce jour.

Les alternatives au DPM

Le modèle qui fait actuellement concurrence au DPM est le modèle *Conséquences*, *Normes* et *Inaction vs. Action* (CNI, Gawronski et al., 2017). Le modèle CNI a été conçu pour les prises de décision morales mais il s'applique également aux jugements moraux.

Ce modèle repose sur le précepte que les jugements utilitaires sont inférés à partir d'une sensibilité aux conséquences morales (C), tandis que les jugements moraux déontologiques sont basés sur la sensibilité aux normes morales (N). Le modèle fait

intervenir une troisième facette : une préférence générale pour l'inaction ou l'action (I), et ce indépendamment des conséquences et des normes.

Concernant les aspects émotionnels et cognitifs avancés par le DPM, ils ne jouent plus un rôle central où l'un des aspects va supplanter l'autre selon certaines conditions. Dans ce modèle, ces deux aspects influencent chacun des facettes spécifiques de façon conjointe (les conséquences, les normes ou la préférence pour l'action-inaction).

En effet, Gawronski et al. (2017) suggèrent que les aspects émotionnels (l'implication émotionnelle) vont avoir tendance à réduire la sensibilité aux normes morales et augmenter la préférence pour l'inaction. Quant aux processus cognitifs, ils influencent uniquement la préférence pour l'action ou l'inaction. Par exemple, en cas de saturation cognitive, les individus vont avoir une préférence plus marquée pour l'inaction (Gawronski et al., 2017). Les auteurs interprètent ce résultat comme suit : moins les individus ont la capacité d'analyser comme il se doit les tenants et aboutissants d'un dilemme qui les forcent à réaliser un choix, plus ils vont vouloir se désengager d'une prise de décision.

Ce modèle, bien que prometteur (étant donné qu'il offre un éclairage plus précis sur le rôle des processus émotionnels et cognitifs), présente une limite majeure. La principale difficulté réside dans la rigidité d'utilisation du modèle. De fait, pour tester le pouvoir explicatif et prédictif du modèle, il est nécessaire (à ce stade) d'utiliser un matériel spécifique faisant intervenir les trois aspects du modèle de façon séparée. En d'autres termes, pour un même dilemme, trois variantes (elles-mêmes subdivisées en deux) ont été créées testant chacune l'aspect *Conséquence*, *Normes*, ou *Inaction*. En comparaison, le DPM a permis une classification (peut-être quelque peu triviale) de tous les dilemmes de type trolley en amont (i.e., Impersonnel vs. Personnel). Dès lors, quel que soit le dilemme utilisé, il existe une nomenclature classant *a priori* le dilemme et indiquant les processus qui sont davantage engagés dans ce dilemme. Ainsi, dans sa forme actuelle, le modèle CNI repose sur un matériel conséquent (i.e., six variantes par dilemme). Toutefois, il apporte une approche théorique plus riche et nuancée que le DPM.

Un modèle complémentaire issu des études en économie :

Comme nous l'avons vu, il existe un modèle, le DPM, qui explique les jugements et choix moraux de façon limitée, et un modèle (plus prometteur) qui nécessite une manipulation conséquente des scénarios pour étudier les jugements et choix moraux.

Au vu des différentes limites énumérées précédemment, ces modèles apparaissent suboptimaux lorsque l'on souhaite les utiliser comme outil prédicteur des choix ou jugements moraux. Certes, ils offrent une explication théorique probante mais il apparait nécessaire qu'une approche alternative puisse offrir un meilleur pouvoir de prédiction.

Pour pallier cette difficulté, il apparait opportun de proposer une approche alternative issue des recherches en économie pour évaluer les comportements de clients et leur promptitude à acheter. Cette méthode est basée sur une variante de la technique d'Utilité Multi-Attributs utilisée dans les prises de décision économiques (UMA, Von Winterfeldt & Edwards, 1986). La technique est basiquement focalisée sur une analyse conjointe adaptative (Johnson, 1987) qui, à travers plusieurs étapes, qui amènent les individus à sélectionner une alternative « optimale » (e.g., un produit). Bien que ce type de technique a été employée principalement en économie, elle est applicable à d'autres champs où un décideur doit procéder à la sélection d'une alternative (une option) parmi plusieurs. C'est d'ailleurs la raison pour laquelle ce type d'approche également apparaitrait comme un type de modèle plus pertinent pour expliquer les choix moraux dans les dilemmes de type trolley.

Par souci de clarté, ces étapes seront illustrées via un exemple ne faisant pas intervenir d'aspects moraux. Ensuite, nous le transposerons aux dilemmes de type trolley.

Les fondements des techniques UMA

Tout problème a de nombreuses dimensions que l'on caractérise en attributs (Jansen, 2011). Au cours d'un processus de prise de décision, le décideur va considérer un ensemble d'alternatives (i.e., les options possibles ou les différents choix qui s'offrent à lui) en considérant tous les attributs inhérents au problème. Dans cette analyse, le but principal du décideur est qu'il puisse dégager un optimum (i.e., une alternative qui présente les meilleurs attributs ou, tout du moins, les moins dommageables).

Ainsi, l'individu va devoir sélectionner des attributs significatifs et les pondérer selon leur importance relative. De ce fait, les attributs qui seront jugés comme étant plus importants auront davantage de poids dans l'émergence de l'optimum.

L'idée principale à travers l'usage de ce type de technique dans des problèmes de type trolley est de cartographier les attributs communs que l'on retrouve chez la majorité des individus. Ensuite, cela permettra de déterminer la manière dont les attributs sont pondérés en fonction du type de dilemme. Sur cette base, nous pourrons plus aisément déterminer à quel moment du processus de prise de décision interviennent des facteurs de personnalité, émotionnels ou cognitifs (ainsi que leur poids relatif).

Von Winterfeldt & Edwards (1986) ont recensé la séquence générale que l'on retrouve dans les techniques UMA :

1. Définir les alternatives et les attributs pertinents du problème ;
2. Evaluer les alternatives ;
3. Assigner une pondération relative aux différents attributs ;
4. Réaliser une évaluation globale des différentes alternatives ;
5. Emergence d'un optimum.

La caractérisation : définition des alternatives et des attributs pertinents :

Cette première phase est certainement la plus complexe. En effet, elle fait appel à un travail d'analyse qui vise à déterminer les attributs centraux communs aux différentes alternatives. Le nombre et la qualité des attributs se doivent d'être *complets* (i.e., tous les aspects du problème doivent être brassés), *opérationnels* (i.e., les attributs doivent jouer un rôle significatif dans l'analyse), *non-redondants* (i.e., chaque attribut doit être unique) et *minimal* (i.e., le nombre d'attributs doit être le plus petit possible) (Keenay & Raiffa, 1976). Cette étape vise un nombre de *k*-attributs ; classés selon différents niveaux. Ce nombre d'attributs est, dans l'absolu, illimité, mais il sera préférable de le limiter à $k \leq 5$ caractéristiques les plus saillantes et les plus importantes pour l'individu.

Dans l'absolu, von Winterfeldt & Edwards (1986) recommandent que les attributs puissent s'évaluer selon une échelle linéaire ou ordinale[8] (que l'attribut soit quantitatif ou qualitatif)

Prenons l'exemple d'un individu qui reçoit trois propositions d'emploi, chacune ayant ses particularités. De plus, il s'agit d'un choix important pour cet individu car il va occuper cet individu pour une période définie et ce dernier ne va pas pouvoir cumuler les emplois. Il s'agit donc d'alternatives mutuellement exclusives. Dans cet exemple, il faut définir les attributs principaux communs aux trois emplois : c'est une phase de collecte d'informations et de synthèse.

Dans le cadre de ces trois emplois, les attributs centraux sont sur base des informations disponibles : (1) les horaires, (2) le salaire, (3) la fonction qui met le plus en avant les compétences de l'individu et (4) la proximité géographique (voir Tableau 2 ci-après décrivant les particularités de chaque emploi[9]).

Notez qu'il existe d'autres manières de réaliser la cotation et la pondération. L'adaptation de la technique UMA présentée dans ce chapitre vise principalement à ce qu'elle puisse être facilement implémentée dans un contexte expérimental (e.g., facilement utilisée par des participants).

[8] Selon certaines approches UMA, il est en effet possible que certains attributs adoptent une fonction concave ou convexe : il y a un optimum et, au-delà ou au-dessous, la préférence est dévaluée. Par exemple, l'optimum du nombre d'heures de travail pour un individu est 35 heures. Dès lors, un emploi qui offrira 22 heures ou 45 heures de travail sera côté plus faiblement qu'un emploi qui se rapproche des 35 heures (pour davantage de détails, voir Jansen, 2011). Toutefois, via l'alternative UMA que nous proposons, ce type de fonction n'est pas rencontrée.

[9] D'autres caractéristiques comme le type de contrat pourrait également être pris en considération mais, pour faciliter la démonstration du modèle, nous allons considérer que les trois emplois proposent un contrat à durée déterminée d'un an (ce qui équivaut à attribuer 1 point à chaque option).

Tableau 2. Etape 1 : définition des alternatives et attributs

	Alternatives		
Attributs	**Emploi 1**	**Emploi 2**	**Emploi 3**
Horaires	Nombre d'h/sem	Nombre d'h/sem	Nombre d'h/sem
Salaire net/mois	€ net/mois	€ net/mois	€ net/mois
Fonction (perception subjective)	Faiblement / Moyennement / Fortement pertinente	Faiblement / Moyennement / Fortement pertinente	Faiblement / Moyennement / Fortement pertinente
Proximité géographique en km	Nombre de km	Nombre de km	Nombre de km

La cotation : évaluation conjointe des attributs :

La seconde étape vise à coter les préférences des différentes caractéristiques recensées (de 1 à 3, car il y a 3 options/emplois). En effet, nous supposerons que le décideur a une préférence pour un horaire « plus léger », un salaire plus attrayant, un emploi qui présente la meilleure pertinence entre la fonction proposée et ses compétences, et enfin, pour l'emploi le plus proche de son domicile. Dans cette étape, si deux caractéristiques sont jugées égales, on leur attribuera la même cotation. Une fois la cotation terminée, on réalise un simple comptage en additionnant les points attribués pour chaque option (une option correspond à un emploi). Le Tableau 3 illustre cette étape de cotation et de comptage.

Toutefois, intervient (potentiellement) un autre facteur important dans cette étape : la pondération. En effet, jusqu'à présent, nous sommes partis du principe que tous les attributs définis étaient équivalents. En d'autres termes, aucune de ces caractéristiques définies n'est apparue comme plus importante que les autres. Néanmoins, si la proximité géographique présente une importance prépondérante pour notre individu, il pourra déterminer un nombre de points additionnels à soustraire au comptage (e.g., -2 points à l'alternative cotée 1). Ce point serait notamment déterminé par les préférences

« passionnées » de l'individu mais aussi par les émotions qu'il va ressentir au cours de son évaluation (les processus émotionnels).

Tableau 3. Etape 2, évaluation des attributs

Attributs	Emploi 1	#	Emploi 2	#	Emploi 3	#
Horaires	37h/sem	1	38h/sem	2	38h/sem	2
Salaire net/mois	2000	3	2100	2	2300	1
Fonction (perception subjective)	Moyennement Pertinente	2	Faiblement pertinente	3	Fortement pertinente	1
Proximité géographique en km	5km	1	17km	2	17km	2
Total		7		9		6

Le tableau reprend les différentes informations de chaque emploi et la cotation attribuée par le décideur.

L'émergence de l'optimum ; le pré-choix et la décision finale :

La dernière étape vise à réaliser un choix (i.e., sélectionner une option unique) qui a été optimisé par les deux précédentes étapes. En l'occurrence, l'alternative qui présente le nombre de points le plus bas (i.e., l'utilité multi-attributs la plus optimale) sera privilégiée, réanalysée et, enfin, l'individu décidera s'il opte ou non pour cette alternative.

Dans notre exemple, on constate que l'emploi numéro 3 apparait comme l'alternative la plus optimale pour notre individu (il comptabilise 6 points), bien que très proche de l'emploi 1 qui comptabilise 7 points. Ainsi, le pré-choix s'oriente vers l'emploi 3 ; il s'agit de l'option saillante, l'optimum. Néanmoins, si l'individu avait donné une importance prépondérante à la proximité géographique, l'emploi 1 s'avérerait la meilleure option. Ce type de modèle de prise de décision vise à maximiser l'utilité (e.g., le produit qui va être consommé ou acheté ou celui pour lequel le décideur va opter). Il présente un

avantage de taille, celui d'intégrer et d'analyser conjointement les différents attributs des multiples alternatives d'un problème.

Limites inhérentes aux techniques UMA

Les trois principales limites ou difficultés du modèle sont : (1) les variables inconnues (*unknown variables*), (2) les variables mal évaluées (*misjudged variables*) et (3) les variables omises (*omitted variables*)[10].

Les *variables inconnues* sont des facteurs significatifs qui ne sont pas accessibles à l'individu qui prend une décision. En d'autres termes, ce sont des facteurs qui sortent de son champ de maîtrise et qui ne sont pas anticipables (i.e., elles ne sont pas accessibles lors de la collecte d'informations). Toutefois, si elles avaient été intégrées dans le processus de prise de décision, elles auraient pu influencer significativement son choix. Pour l'exemple des trois emplois, une variable inconnue serait des retards importants dans le versement des salaires de la part de l'employeur.

Pour les *variables mal évaluées*, il s'agit de facteurs significatifs qui ont été évalués erronément par le décideur. Dans notre illustration, une variable mal évaluée serait la proximité géographique. Effectivement, il aurait été plus pertinent d'évaluer le temps de trajet moyen entre le domicile et le lieu de travail plutôt que la distance géographique relative : un emploi peut être proche géographiquement du domicile d'un individu mais difficile d'accès ou surchargé en trafic ; ce qui peut donc fausser l'étape de cotation.

Enfin, les *variables omises* sont des variables significatives qui auraient pu être prises en considération par le décideur mais qui ont simplement été négligées ou oubliées dans l'étape de caractérisation (alors que l'information était disponible). Par exemple, une variable omise serait les possibilités d'évolution dans l'entreprise dans un futur proche.

Une dernière limite de ce type de modèle est le caractère *a priori* arbitraire des étapes de caractérisation et de cotation. Toutefois, comme nous le verrons ultérieurement, certains patterns interindividuels semblent émerger, ce qui présume un ciblage et une évaluation approximativement similaires des mêmes attributs chez les individus.

[10] Une variable correspond à un attribut.

Application des modèles de type analyse conjointe multi-attributs aux problèmes du trolley

Bien qu'il s'agisse de modèle de prise de décision à visée économique, ils peuvent s'appliquer aux dilemmes moraux de type trolley. Pour rappel, les techniques UMA, comme leur nom l'indique, repose sur la maximisation de l'utilité en prenant en compte de multiples attributs. Dans un contexte économique, l'utilité correspond à la propension à acheter ou consommer un produit, mais cela permet également de déterminer la préférence de l'individu pour un produit. Dans un contexte moral, l'utilité correspondrait à la propension à agir ou à déterminer la préférence de l'individu pour une alternative spécifique (qui peut être la moins préjudiciable).

Ces problèmes sont majoritairement réduits à deux alternatives dans les prises de décision morales : sacrifier les cinq pour sauver l'individu isolé ou sacrifier l'individu isolé pour sauver les cinq. En reprenant le dilemme du trolley de Foot (1967), nous avons donc deux voies où, d'un côté, se trouvent cinq ouvriers et, de l'autre, un seul. Le décideur doit choisir vers quelle voie il orientera le train qui tuera inéluctablement le ou les ouvrier(s) qui se trouveront sur son chemin.

L'étape 1 du modèle vise à sélectionner les attributs importants pour les deux options en vue d'une prise de décision. Pour ce faire, reprenons différents éléments théoriques issus des modèles abordés précédemment. Le premier attribut serait de privilégier l'option qui va sauver le plus de personnes (que l'on nommera « l'optimisation »), concept que l'on doit à l'utilitarisme (i.e., maximiser l'utilité et centrer son choix sur les conséquences de ce dernier plutôt que l'action en tant que telle). Un deuxième attribut serait les normes morales violées dans l'une des options ; attribut issu de l'approche déontologique que l'on retrouve également dans le modèle CNI (Gawronski et al., 2017). Un autre volet d'attributs fait intervenir des aspects post-décisionnels comme l'impact de la décision sur l'image personnelle (est-ce que l'action est en phase avec mes valeurs ?) et l'impact de la décision sur l'image sociale (i.e., comment les autres personnes vont juger ma décision ?). Svenson (1992) soutient que les aspects post-décisionnels sont capitaux dans les prises de décision. Dès lors, l'impact sur l'image personnelle et sociale sont deux caractéristiques post-décisionnelles importantes.

Dans notre exemple, nous allons nous baser sur le fait que le dilemme du trolley de Foot n'implique pas de normes morales violées. Pour rappel, on considère que ce type de dilemme est *impersonnel*, notamment parce qu'il n'implique pas de violer des interdits moraux (contrairement aux dilemmes de type *personnels* où il faut réaliser une action intentionnelle qui vise à sacrifier un ou cinq individus). L'impact sur l'image personnelle dépendra de l'évaluateur et l'impact sur l'image sociale aussi. Le Tableau 4 illustre cette étape.

Tableau 4. Etape 1 : La caractérisation

Caractéristiques	**Sauver les cinq**	**Sauver l'individu isolé**
Optimisation	Oui/Non	Oui/Non
Normes morales violées	Oui/Non	Oui/Non
Impact sur l'image personnelle	+/-/0	+/-/0
Impact sur l'image sociale	+/-/0	+/-/0

La deuxième étape consistera à coter les différents attributs (de 1 à 2, étant donné qu'il y a deux options). Nous allons considérer que : (1) le fait de sauver cinq personnes et d'en sacrifier une à un impact négatif sur l'image personnelle, comme le fait d'en sauver une et d'en sacrifier cinq ; (2) les membres de sa sphère sociale jugeront plus propice de sauver cinq personnes par rapport à une seule (Cf. Tableau 5 illustrant cette étape).

Tableau 5. Etape 2 : La cotation et le comptage

Caractéristiques	**Sauver les cinq**	**#**	**Sauver l'individu isolé**	**#**
Optimisation	Oui	1	Non	2
Normes morales violées	Non	1	Non	1
Impact sur l'image personnelle	+	1	+	1
Impact sur l'image sociale	+	1	-	2
Total		4		6

La dernière étape nous oriente vers un optimum : l'alternative utilitariste (i.e., « sauver les cinq »). Comme nous l'avions mentionné, si notre individu attribue une importance prépondérante à l'optimisation, nous retirerons 2 points au total de l'option qui a reçu la cote la plus basse à cette caractéristique (ici, sauver les cinq) ; ce qui aura pour effet de rendre l'alternative utilitariste encore plus saillante.

Dans le cadre des dilemmes de type trolley, d'autres facteurs pourraient être pris en considération : Par exemple :

- Le degré de dérobade face au problème (i.e., le degré d'échappatoire dans une situation complexe ; dans le cadre des dilemmes de type trolley). Dans les dilemmes de type trolley, la dérobade est généralement contrôlée, donc nulle ;
- Les regrets que pourraient générer un choix spécifique ;
- La qualité (positive ou négative) des protagonistes (si elle est décrite dans le scénario) ;
- La certitude du résultat (e.g., le fait que dévier le train sur l'individu isolé sauve effectivement les cinq).

Ces points clés ont été repris des différents travaux suivants : on retrouve le concept de dérobade de Batson (Batson et al., 1981 ; Batson, 1987 ; 2011) qui rejoint la préférence pour l'inaction vs. action propre au modèle CNI (Gawronski et al., 2017) ; l'incertitude du résultat de Kahneman et Tversky (Kahneman, 2013 ; Kahneman & Tversky, 1982) montrant que le degré de certitude ou d'incertitude d'un résultat influence les décisions des individus ; les deux critères centraux du modèle CNI : utilitarisme et violation des normes morales ; enfin, concernant la qualité des individus, on sait par exemple que le degré d'intimité entre un(e) participant(e) et un(e) protagoniste peut influencer significativement la prise de décision. En effet, si l'un des protagonistes est une personne proche, la décision morale est facilitée car le(la) participant(e) éprouvera moins de réponses et de conflits émotionnels (Zhan et al., 2018) : il(elle) choisira de sauver ce proche. Un autre exemple concernant la qualité des protagonistes serait la responsabilité causale de ces derniers (Bartels et al., 2015) : si le décideur perçoit un potentiel élément de faute ou de

responsabilité chez l'un des protagonistes, cela peut influencer significativement sa prise de décision[11].

Il est postulé que les individus appliquent ce modèle de manière intuitive en ciblant et/ou en pondérant des attributs limités. Les intérêts principaux des techniques UMA sont : (1) la systématisation du processus de prise de décision, (2) de cerner les attributs centraux pris en considération dans un domaine précis de prise de décision et, enfin, (3) potentiellement d'aiguiller les individus dans leurs prises de décision. L'alternative proposée ci-avant reprend les deux aspects centraux du modèle DPM en montrant que l'analyse conjointe multi-attributs est un processus essentiellement réflexif où les processus émotionnels ou les préférences du décideur (i.e., les émotions évoquées par la situation, les attraits personnels ou les dispositions émotionnelles) vont diriger le focus attentionnel vers un ou des attributs spécifiques qui vont devenir prépondérants par rapport aux autres (e.g., le décideur va se limiter à évaluer à un seul attribut).

Ceci révoque l'idée que les jugements déontologiques sont plutôt « *model-free* » tandis que les jugements utilitaires sont plutôt « *model-based* » (Miller & Cushman, 2013 ; Patil et al., 2021). En effet, les deux types de prise de décision ou jugement peuvent être *model-free* ou *model-based*. Par exemple, si l'attribut optimisation demeure central pour l'individu, il ne se lancera pas dans une collecte d'information et ne procèdera pas à une analyse conjointe multi-attributs. En quelque sorte, sa préférence « passionnée » aura phagocyté ses ressources attentionnelles et il est peu probable qu'il se lance dans un processus plus élaboré et couteux en ressources cognitives. Ce choix utilitariste apparaitra donc *model-free* (i.e., modèle peu élaboré et limité en termes d'attributs). Toutefois, ce même individu aurait tout à fait pu être mu par une préférence passionnée pour le respect des normes morales et mettre l'accent sur cet attribut unique. En découlerait donc, un choix déontologique que l'on pourrait qualifier également de *model-free.*

À la lumière des résultats obtenus dans nos trois études (i.e., approximativement les mêmes patterns de choix sont observés dans les dilemmes avec des échantillons indépendants), il est spéculé qu'il existe des patterns (intuitifs) interindividuels de caractérisation. Autrement dit, il y a des attributs récurrents qui sont évaluées de façon similaire chez la plupart des individus et il est probable que les états émotionnels des

[11] Le lecteur est invité à se référer aux Annexes F & G qui illustrent l'application des modèles d'analyse conjointe multi-attributs à d'autres dilemmes de type trolley.

individus ou leurs préférences vont focaliser sur le même nombre limité d'attributs. C'est d'ailleurs probablement la raison pour laquelle les effets des facteurs émotionnels et cognitifs sont faibles sur les prises de décision morales (i.e., l'*output*). Effectivement, ces facteurs agiraient plus précocement dans le processus de prise de décision : dans les étapes de définition des attributs et d'évaluation, plutôt que dans la décision en elle-même.

À ce stade, l'application de ce type de modèles basés sur l'analyse conjointe d'attributs dans des dilemmes de type trolley est spéculatif et s'insère comme une alternative potentielle permettant d'expliquer voire de prédire les jugements et choix moraux. De par sa simplicité d'utilisation, l'intérêt principal des techniques UMA est qu'elles visent à répondre au décalage observé entre les prédictions des modèles théoriques existants et les données empiriques.

Conclusions

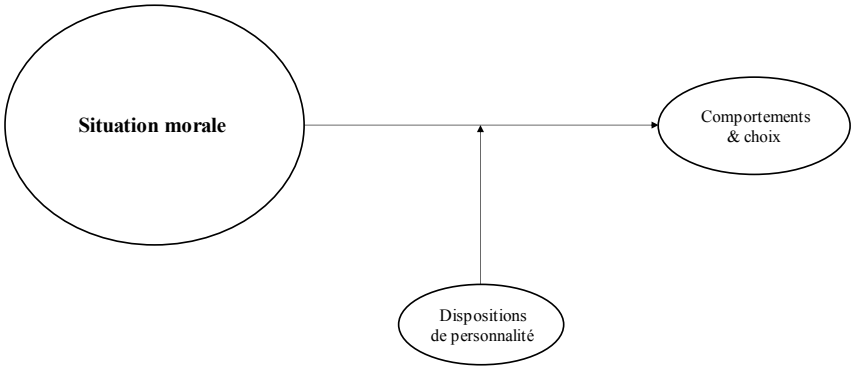
Cette thèse de doctorat a permis de mettre en lumière différents points importants concernant les prises de décision morales dans les dilemmes de type trolley.

Premièrement, nous avons pu démontrer un effet causal majeur de la perspective du participant sur les prises de décision. Notamment, la perspective B (où le(a) participant(e) doit choisir entre lui(elle) et les cinq autres protagonistes) est apparue comme particulièrement déterminante dans les changements de choix. Comme l'ont montré nos résultats, on retrouve des patterns approximativement similaires entre les perspectives A (i.e., là où le(a) participant(e) incarne le *bystander*) et C (i.e., lorsque le(a) participant(e) incarne un des protagonistes parmi les cinq). Ceci démontre que les individus, lorsqu'ils doivent prendre une décision entre la minorité et la majorité, vont plus souvent se mettre à la place de la majorité. Toutefois, lorsqu'ils incarnent eux-mêmes la minorité, la prise de décision change drastiquement (i.e., on retrouve un pattern de 50-50 entre les choix utilitaires et déontologiques). Cette différence pourrait s'expliquer par deux facteurs : (1) soit les individus font intervenir davantage d'attributs dans cette perspective, comparativement aux deux autres, ce qui les conduit à des choix plus partagés ; (2) soit la tourmente générée par cette perspective les conduit à pondérer de manière variable des attributs spécifiques et limités.

Deuxièmement, nous avons pu démontrer un second effet causal lié au type de dilemme. Bien que cet effet fût déjà connu entre dilemmes « classiques » (i.e., impersonnels et personnels), les variations entre des dilemmes de type trolley de la vie de tous les jours (e.g., les dilemmes de l'entreprise et du départ anticipé) sont également apparues comme majeures. Pour illustration, dans la perspective A, nous avons observé des changements de 90 à 10 % de choix orientés vers le groupe entre ces deux types de dilemme (90% de choix utilitaires pour le dilemme de l'entreprise et 10% de choix utilitaires pour le dilemme du départ anticipé). Au regard de ces changements significatifs entre les types de dilemmes, l'interprétation la plus probable est que chaque situation morale, selon les caractéristiques qu'elles comportent (et le fait qu'incarner une perspective particulière puisse également modifier la perception des caractéristiques de la situation), est susceptible d'engendrer un pattern de prise de décision spécifique. Ceci rejoint les fondements théoriques de la psychologie sociale montrant que ce sont davantage les

caractéristiques situationnelles qui prédisent et expliquent les comportements que les dispositions personnelles des individus. Tout comme les situations de vie ont un impact important sur les facettes de la personnalité, comme nous l'avons montré dans une étude (Nasello et al., 2023b), les situations morales ont ce même impact sur les choix des individus. Dès lors, étant donné que l'impact des facteurs ou dispositions de personnalité est en général assez faible sur les choix moraux, ces derniers apparaissent plutôt comme un modérateur[12] de l'influence causée par la situation morale sur l'individu (i.e., sur ses comportements ou ses choix), comme illustré dans le Tableau 6 ci-dessous.

Tableau 6 : illustration du rôle modérateur des dispositions de personnalité



Dans la continuité du point précédent, il n'existe probablement pas de lois morales universelles permettant de prédire avec une certitude absolue la manière dont les individus prennent des décisions. En rapport à nos résultats, nous avons fréquemment observé des individus qui optaient pour l'option utilitaire dans certaines perspectives et échangeaient avec l'option déontologique dans d'autres perspectives ; il y a donc une propension marquée à la dérogation de principes moraux. L'interprétation donnée est que chaque situation morale possède ses propres aspects situationnels causaux qui vont catalyser des décisions spécifiques. L'hypothèse soutenue est que les conduites et choix moraux sont essentiellement syncrétiques et paradoxaux. De ce fait, l'individu va tenter de justifier logiquement des comportements ou des choix qui reposent sur des doctrines opposées pour réduire la dissonance cognitive éprouvée et le manque de constance. Par exemple, un

[12] Ce point important illustre une prépondérance du rôle des facteurs situationnels sur la manière d'agir des individus en comparaison aux dispositions ou traits de personnalité. Toutefois, l'individu reste un acteur significatif. Cette formulation est davantage à concevoir de cette façon : les facteurs situationnels vont diriger la majorité des individus à agir ou à opter pour des décisions particulières, mais certaines dispositions individuelles vont rompre avec ce newtonianisme.

individu pourrait juger approprié de réaliser lui-même une manœuvre létale sur un patient qui présente une pathologie bénigne pour sauver cinq autres patients mourants, et juger un cas similaire comme inapproprié. Autrement dit, même si un individu juge inappropriée une action réalisée par un tiers, il trouvera des justifications logiques et morales pour expliquer son jugement si lui-même est impliqué dans une situation similaire et qu'il juge approprié de réaliser la manœuvre létale.

Troisièmement, nous avons mis en évidence que l'empathie jouait un rôle significatif mais faible dans les prises de décision morales, uniquement dans certaines perspectives. Ceci rejoint amplement les déclarations de Jean Decety et affaiblit le Dual-Process Model (Greene et al., 2001 ; 2004). En effet, le rôle de l'empathie, voire des émotions, est nettement surestimé dans les prises de décision morales et les jugements moraux, comme nous l'avons démontré dans notre méta-analyse et revue systématique (Nasello & Triffaux 2023b). Premièrement, seules certaines facettes de l'empathie apparaissent comme significativement associées aux différents domaines de la morale (i.e., l'empathie affective et non l'empathie cognitive). Deuxièmement, les tailles de l'effet sont relativement faibles. Pour expliquer cette taille d'effet, l'interprétation avancée est que l'empathie jouerait un rôle certainement plus prépondérant dans des aspects antérieurs à la prise de décision, notamment dans les phases de caractérisation et d'évaluation des attributs, comme développé dans la présentation des techniques UMA ci-avant.

En mettant en lien nos travaux sur l'empathie (notamment le développement du modèle d'empathie : Nasello & Triffaux, 2023a) avec ceux sur les dilemmes de type trolley, on peut comprendre de manière plus précise comment ce facteur intervient dans certaines étapes de la prise de décision. En effet, deux voies attirent notre attention dans le processus empathique : celle des attitudes froides et de l'indifférence et celle de la submersion. La voie de la submersion émotionnelle, mise en lien avec les modèles d'analyse conjointe multi-attributs, expliquerait le processus de pondération d'attributs spécifiques : l'individu submergé par des émotions fortes va cibler un nombre limité d'aspects centraux qui vont catalyser sa décision finale. Cette voie est susceptible d'engendrer une focalisation sur un attribut particulier qui deviendra alors central et occultera ou limitera tout autre attribut (cet attribut pourrait autant être déontologique qu'utilitaire). Concernant la voie des attitudes froides et de l'indifférence (qui serait associée aux traits antisociaux et psychopathiques), l'individu serait plus enclin à réaliser un processus de définition des attributs élaboré.

Toutefois, cette prédisposition empathique pourrait l'amener à pondérer davantage certains attributs (e.g., l'optimisation au détriment de la violation des normes morales).

Quatrièmement, les traits borderline sont apparus comme des prédicteurs significatifs des choix moraux uniquement dans un large échantillon issu de la population générale (échantillon comprenant des hommes et des femmes). Ces traits étaient tantôt prédicteurs des choix orientés vers le groupe dans le dilemme de l'entreprise (perspective B), tantôt prédicteurs de ces choix dans le dilemme du chirurgien (perspective C). Les effets étant relativement faibles et ciblés, ils n'ont pas été reproduits dans un échantillon plus restreint de patients présentant un trouble grave de personnalité borderline (comprenant uniquement des femmes). De plus, contrairement aux prédictions réalisées à partir du DPM, ces patientes n'ont pas présenté de différences significatives dans leurs prises de décision par rapport aux volontaires sains présentant des traits borderline faibles et élevés. Au contraire, des effets opposés aux prédictions avaient tendance à se dessiner : les patientes borderline et les contrôles avec des traits borderline élevés réalisaient davantage de choix utilitaires par rapport aux contrôles ayant des traits borderline faibles. De ce fait, nos résultats et conclusions rejoignent le constat émis par de nombreux auteurs soulignant l'incomplétude et le manque de validité du DPM (Smillie et al., 2020 ; Oudman et al., 2021 ; Duke and Bègue, 2015 ; Horne and Powell, 2016). L'auteur principal du modèle, Joshua Greene, a d'ailleurs souligné récemment qu'une révision du modèle devait être réalisée (Greene, 2023). Toutefois, depuis la création du DPM, les modèles alternatifs permettant de prédire et d'expliquer les choix et jugements moraux manquent à l'appel. Un seul modèle concurrent au DPM a été proposé, le modèle CNI (Conséquences, Normes morales, Inaction vs. Action ; Gawronski et al., 2017). Ce modèle, bien qu'il combine les approches déontologiques et utilitaristes sans les opposer, s'étudie dans un cadre d'expérimentation strict qui nécessite de faire varier les dilemmes dans des conditions spécifiques. Malheureusement, cette rigidité d'utilisation contraint le modèle à un champ d'application restreint, ce qui permet difficilement d'expliquer des résultats qui sortent de son cadre expérimental.

En combinant différentes approches théoriques, les techniques d'analyse conjointe multi-attributs pourraient expliquer et prédire de façon plus pertinente les choix moraux dans les dilemmes de type trolley. Par ailleurs, certaines limites sont à prendre en considération ; notamment, les étapes de caractérisation et d'évaluation des attributs sont

sujettes à une variabilité interindividuelle importante en raison de leur caractère arbitraire. Néanmoins, l'investigation d'un domaine de prise de décision particulier à travers les techniques de type UMA permettra de définir les attributs communs et centraux qui sont évalués par les individus. En définissant ces attributs centraux, leur degré de présence et la manière dont les individus sont susceptibles de les pondérer, nous obtiendrons un modèle offrant un pouvoir de prédiction plus performant.

Cinquièmement, notre étude clinique a pu mettre en évidence les domaines clés qui distinguaient les personnes diagnostiquées d'un trouble grave de personnalité borderline des personnes contrôles faibles et élevés en traits borderline[13]. Il est apparu que les patientes présentaient plus de comportements négatifs, d'affects de colère dysfonctionnels, de profond découragement, de perturbations subjectives de l'identité, de perturbations comportementales (impulsivité marquée) et affectives, de fragilité, d'oppositionalité et de précipitation (i.e., manque de délibération). On constate également une propension graduelle à l'anxiété et à la dépression chez les individus diagnostiqués et les contrôles avec des scores élevés en traits borderlines. En d'autres termes, plus les individus présentent des traits borderline élevés plus ils apparaissent vulnérables à l'anxiété et à la dépression comparativement à des individus contrôles ayant des traits borderlines faibles.

Enfin, nous avons pu montrer que les personnes borderline présentaient des styles de prise de décision différents par rapport aux individus contrôles. En effet, les patientes borderline présentaient des styles de décision plus évitants par rapport aux individus contrôles présentant des traits borderline faibles[14], mais surtout moins rationnels en comparaison aux individus contrôles présentant des traits borderline faibles et élevés. Le style de décision « évitant » pourrait être associé aux comportements d'évitement des abandons ou des expériences émotionnelles intenses des personnes borderline, et les scores faibles à la facette rationnelle pourraient être associés à des comportements à risque et impulsifs, aux automutilations ou aux tentatives de suicide. Au regard de ces résultats, les techniques UMA pourrait constituer un matériel clinique prometteur pour les patients éprouvant des difficultés à réaliser des choix au quotidien. Effectivement, les techniques d'analyse conjointe multi-attributs étant relativement simple d'utilisation permettraient de

[13] Note importante : seules des femmes ont été testées dans notre étude. Ces résultats ne sont donc pas généralisables aux hommes qui présenteraient un trouble de personnalité borderline.
[14] Pour rappel, les patientes borderline et les contrôles aux traits borderline élevés avaient des scores comparables au style de décision « évitant », mais ces deux groupes présentaient des scores qui différaient significativement des contrôles aux traits borderline faibles.

fournir une méthode de prise de décision à systématiser ; ce qui pourrait aiguiller les patients à réaliser des choix plus fonctionnels et bénéfiques. *A posteriori*, elles offriraient également une base clinique objectivant les différences observées entre la définition d'un optimum et le choix réalisé par le(la) patient(e). Par exemple, lorsqu'un(e) thérapeute réalise un travail d'analyse sur les choix opérés par un(e) patient(e) dans une situation donnée, cela permettrait de comprendre les mécanismes ayant été mis en place par ce(tte) dernier(ière) au moment de la prise de décision.

Les implications sociétales

Les dilemmes de type trolley offrent un large champ d'application. De plus, l'utilisation de ces dilemmes dans des situations de la vie quotidienne en proposant aux participants d'incarner différents protagonistes permet de proposer une extension significative à ce champ d'application. Nous allons illustrer dans cette section les principaux secteurs pouvant mettre en évidence des dilemmes de type trolley.

Le secteur médical, notamment, regorge de situations illustrant des dilemmes de type trolley. En effet, la pandémie COVID-19 a fait rencontrer à l'ensemble de la population des situations inédites. Par exemple, les dirigeants et experts ont été confrontés aux questions suivantes : doit-on vacciner une petite portion de la population vulnérable ou l'ensemble de la population ? Est-il préférable de mettre en quarantaine l'ensemble de la population ou uniquement les individus diagnostiqués ? Mais également, dans une situation où un personnel soignant doit faire face à des ressources limitées, est-il plus judicieux d'allouer davantage de ressources médicales pour sauver un enfant et moins de ressources pour une personne âgée, ou l'inverse ?

Un autre secteur qui bénéficie fortement des études basées sur les dilemmes de type trolley sont les systèmes de déplacements aéronautiques et terrestres ainsi que les systèmes de déplacements autonomes. En effet, de nombreux questionnements émergent face au développement de ce type de véhicules. Par exemple, pour un véhicule autonome qui est face à un accident imminent, doit-il sauver un groupe d'individus ou sauver son passager (i.e., l'acheteur ou le consommateur dudit véhicule) ? Pour un pilote d'avion face à une panne majeure de son engin qui se dirige sur une ville, faut-il qu'il choisisse de crasher son avion dans un endroit isolé (et donc qu'il se sacrifie avec ses passagers) ou qu'il tente un atterrissage sur cette ville (avec le risque d'engendrer plus de décès) ?

Dans le domaine nautique, un fait d'actualité a illustré un dilemme de type trolley (Cf., Struett, 2023[15]) : en Floride (USA), une dame avec deux enfants est partie en mer faire du parachute ascensionnel (un parachute est accroché à un bateau et la vitesse de ce dernier fait décoller le parachute). Ce jour-là, les conditions météorologiques n'étaient pas

[15] Le lien vidéo suivant illustre également cette affaire : https://www.youtube.com/watch?v=ee9X_gF4UwY.

optimales et des rafales de vent majeures sont apparues une fois le bateau en mer et le parachute déployé avec cette femme et les deux enfants. Pris dans le parachute, le vent fut tellement fort qu'il entraînait le bateau dans la direction opposée au déplacement, et le capitaine prit la décision de couper la corde qui reliait le bateau au parachute pour sauver son équipage et les personnes à son bord (notamment le mari de la femme). Après avoir été entraîné par les vents pendant plusieurs kilomètres, le parachute a fini par percuter un pont, tuant la femme et blessant sévèrement les enfants qui ont fini par s'en sortir miraculeusement. Un procès intenté par le mari à l'encontre de la compagnie nautique est en cours et celui-ci soulèvera certainement un questionnement (moral) autour de la décision du capitaine qui partagera l'audience.

D'autres domaines d'application des dilemmes de type trolley sont, par exemple :

- Les essais cliniques : est-il moralement permissible de sacrifier un petit échantillon de la population pour tester un traitement en vue de l'utiliser pour « sauver » un échantillon plus large d'individus ? Peut-on sacrifier un nombre important d'animaux en vue de sauver une majorité d'êtres humains ? Caviola et al. (2021) ont d'ailleurs démontré selon quel degré il nous apparait moralement permissible de sacrifier certains animaux particuliers (e.g., des cochons) en comparaison à d'autres espèces ou à des humains. Ils concluent en montrant que nous attribuons aux animaux une qualité morale similaires aux objets inanimés. En effet, nous pouvons les posséder, les utiliser comme outil, les dépourvoir de leur autonomie et les considérer comme interchangeables. Toutefois, les animaux bénéficient, à l'instar des objets, d'une *relative* protection morale (sans égaler celle à l'égard des humains) et certains d'entre eux vont même bénéficier d'une attention plus particulière ;
- L'usage des armements nucléaires : peut-on sacrifier une portion d'une population ennemie (pour la pousser à capituler) en vue de sauver sa propre population[16] ;
- Les carrières professionnelles : les individus privilégient-ils leur carrière au détriment de celle de leurs collègues ? De leur famille ? ;
- Certains symptômes cliniques : certains profils de patients vont réaliser des choix plus individualistes au détriment des proches. D'autres ne vont réaliser que des choix en faveur de leurs proches au détriment d'eux-mêmes ;

[16] Pour autant que le camp adverse ne possède pas le même armement.

- Certains choix de *parenting* ou certains agencements familiaux : un (ou des) parents qui va(vont) privilégier un enfant d'une fratrie au détriment des autres, ou l'inverse ;
- Au niveau juridique : un individu qui prend la décision de tuer intentionnellement un autre individu pour sauver un groupe de personnes, devrait-il être jugé différemment sur le plan pénal ?

Comme illustré, il existe une pléthore de champs d'application des dilemmes de type trolley. Ces derniers nous permettent d'étudier et comprendre la nature humaine tant au niveau de l'adhérence à certains principes moraux qu'au rôle que nous conférons à certaines espèces, mais ils nous éclairent également sur la compréhension de certains troubles. Effectivement, plusieurs études ayant utilisé des dilemmes de type trolley ont permis de comprendre que les individus présentant des traits élevés en psychopathie discernaient parfaitement bien le fait qu'un comportement soit moralement inapproprié. Ces derniers n'en tiennent simplement pas compte dans leur prise de décision (Cima et al., 2010 ; Pletti et al., 2017). Malgré cet éclairage important offert par les dilemmes de type trolley, la principale limite est qu'ils ne prescrivent pas de façon absolue les jugements, les choix ou les conduites morales à adopter ; ils évaluent les tendances générales de ce qui pourrait être moralement acceptable et permettent de déterminer les facteurs qui prédisent ces tendances. L'interprétation de la justesse d'un jugement, d'un choix ou d'une conduite ne repose que sur un consensus le plus souvent *non-généralisable*[17] (car défini au cas-par-cas), lui-même pouvant être limité à une infime portion de la population générale (e.g., par des experts d'un domaine).

En conclusion, les dilemmes de type trolley soulèvent des questionnements moraux aux implications sociétales majeures. Le présent travail a permis d'illustrer et d'élargir le champ d'application des dilemmes de type trolley, de déterminer comment les choix moraux peuvent varier significativement et de définir l'importance relative des facteurs situationnels face aux traits ou dispositions de personnalité (comme l'empathie ou les traits de personnalité borderline).

[17] Ce point n'entre pas en contradiction avec ce qui a été évoqué précédemment. Effectivement, nous pourrions définir les attributs centraux qui interviennent dans les dilemmes de type trolley. Nous pourrions également découvrir les tendances générales des choix qu'opéreraient la plupart des individus. Toutefois, cela ne constituerait pas une ligne directive transcendante qui guiderait les individus dans des conflits moraux.

Ce travail a également permis d'enrichir les connaissances autour du trouble de personnalité borderline, tout en proposant de nouvelles pistes d'intervention thérapeutique. En effet, l'adaptation des techniques d'analyse conjointe multi-attributs avancés dans le cadre de ce travail pourrait avoir de nombreuses applications prometteuses en thérapie. Ils permettraient à des patients en difficulté ou en incapacité de réaliser des choix qui leur sont bénéfiques à adopter une méthodologie qui les aiguilleraient dans leurs décisions du quotidien. De plus, ils offriraient aux thérapeutes la mise en exergue d'éventuelles facettes dysfonctionnelles du processus de prise de décision mis en place par le(a) patient(e) ; ce qui en faciliterait le traitement.

En outre, l'application des techniques UMA aux dilemmes de type trolley offrirait une approche alternative prometteuse pour identifier les attributs centraux qui interviennent dans les prises de décision morales ; tout en identifiant la manière dont les individus pondèrent ces attributs. Ces techniques, ayant pour avantage une simplicité d'emploi, offriraient, à terme, un modèle prédicteur des choix moraux plus performant.

Références

Armstrong, J., Friesdorf, R., & Conway, P. (2019). Clarifying gender differences in moral dilemma judgments: The complementary roles of harm aversion and action aversion. *Social Psychological and Personality Science*, *10*(3), 353–363. https://doi.org/10.1177/1948550618755873

Balash, J., & Falkenbach, D. M. (2018). The ends justify the meanness: An investigation of psychopathic traits and utilitarian moral endorsement. *Personality and Individual Differences*, *127*, 127–132. https://doi.org/10.1016/j.paid.2018.02.009

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral judgment and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making.* Wiley.

Batson, C. D. (1987). Prosocial motivation: Is it ever truly altruistic? *Advances in Experimental Social Psychology*, *20*, 65-122. https://doi.org/10.1016/S0065-2601(08)60412-8

Batson, C. D. (2011). *Altruism in humans*. Oxford University Press.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation?. *Journal of Personality and Social Psychology*, *40*(2), 290.

Batson, C. D., Lishner, D. A., & Stocks, E. L. (2015). The empathy—Altruism hypothesis. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford handbook of prosocial behavior* (pp. 259–281). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399813.013.023

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. https://doi.org/10.1111/spc3.12131

Bentham, J. (1789). *Introduction to the Principles of Morals and Legislation*. London, Payne.

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413–425. https://doi.org/10.2307/1128984

Burghart, M., & Mier, D. (2022). No feelings for me, no feelings for you: A meta-analysis on alexithymia and empathy in psychopathy. *Personality and Individual Differences*, *194*, 111658. https://doi.org/10.1016/j.paid.2022.111658

Choe, S. Y., & Min, K. H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment & Decision Making*, *6*(7), 580–592.

Cima, M., Tonnaer, F., & Hauser, M. D. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience*, *5*(1), 59-67. https://doi.org/10.1093/scan/nsp051

Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A biosocial developmental model of borderline personality: Elaborating and extending Linehan's theory. *Psychological Bulletin*, *135*(3), 495. https://psycnet.apa.org/doi/10.1037/a0015616

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144-153. https://doi.org/10.1177%2F1754073914558466

Curşeu, P. L., & Schruijer, S. G. (2012). Decision styles and rationality: An analysis of the predictive validity of the General Decision-Making Style Inventory. *Educational and Psychological Measurement*, *72*(6), 1053-1062. https://doi.org/10.1177/0013164412448066

D'Ambrosio, F., Olivier, M., Didon, D., & Besche, C. (2009). The basic empathy scale: A French validation of a measure of empathy in youth. *Personality and Individual Differences*, *46*(2), 160–165. https://doi.org/10.1016/j.paid.2008.09.020

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalogue of Selected Documents in Psychology*, *10*, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113. https://psycnet.apa.org/doi/10.1037/0022-3514.44.1.113

de Waal, F., Macedo, S. E., & Ober, J. E. (2006). *Primates and philosophers: How morality evolved*. Princeton University Press.

Decety, J. (2016). Composants, mécanismes, développement et fonctions de l'empathie. *BMC Psychiatry, 13*(1), 1–8.

Decety, J. (2021). Why empathy is not a reliable source of information in moral decision making. *Current Directions in Psychological Science, 30*(5), 425-430. https://doi.org/10.1177%2F09637214211031943

Decety, J., & Cowell, J. M. (2014). Friends or foes: Is empathy necessary for moral behavior? *Perspectives on Psychological Science, 9*(5), 525–537. https://doi.org/10.1177/17456 91614 545130

Decety, J., & Cowell, J. M. (2015). Empathy, justice, and moral behavior. *AJOB Neuroscience, 6*(3), 3-14. https://doi.org/10.1080/21507740.2015.1047055

DeShong, H. L., Mullins-Sweatt, S. N., Miller, J. D., Widiger, T. A., & Lynam, D. R. (2016). Development of a short form of the five-factor borderline inventory. *Assessment, 23*(3), 342–352. https://doi.org/10.1177/1073191115581475

Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition, 134*, 121-127. https://doi.org/10.1016/j.cognition.2014.09.006

Eklund, J. H., & Meranius, M. S. (2021). Toward a consensus on the nature of empathy: A review of reviews. *Patient Education and Counseling*, *104*(2), 300-307. https://doi.org/10.1016/j.pec.2020.08.022

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. In Foot (1977). *Virtues and vices and other essays in moral philosophy*. Oxford: Basil Blackwell, pp.19–32.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of Personality and Social Psychology, 4*(2), 195.

Fumagalli, M., Ferrucci, R.,Mameli, F.,Marceglia, S., Mrakic-Sposta, S., Zago, S., Lucchiari, C., Consonni, D., Nordio, F., Pravettoni, G., & Cappa, S. (2010a). Gender-

related differences in moral judgments. *Cognitive Processing, 11*(3), 219–226. https://doi.org/10.1007/s10339-009-0335-2.

Fumagalli, M., Ferrucci, R.,Mameli, F.,Marceglia, S., Mrakic-Sposta, S.,Zago, S., Lucchiari, C., Consonni, D., Nordio, F., Pravettoni, G., & Cappa, S. (2010a). Gender-related differences in moral judgments. *Cognitive Processing, 11*(3), 219–226. https://doi.org/10.1007/s10339-009-0335-2

Fumagalli, M., Vergari, M., Pasqualetti, P., Marceglia, S., Mameli, F., Ferrucci, R., Mrakic-Sposta, S., Zago, S., Sartori, G., Pravettoni, G., Barbieri, S., Cappa, S., & Priori, A. (2010b). Brain switches utilitarian behavior: Does gender make the difference? *PLoS One, 5*(1), e8865. https://doi.org/10.1371/journal.pone.0008865

Fumagalli, M., Vergari, M., Pasqualetti, P., Marceglia, S., Mameli, F., Ferrucci, R., Mrakic-Sposta, S., Zago, S., Sartori, G., Pravettoni, G., Barbieri, S., Cappa, S., & Priori, A. (2010b). Brain switches utilitarian behavior: Does gender make the difference? *PLoS One, 5*(1), e8865. https://doi.org/10.1371/journal.pone.0008865.

Gao, Y., & Tang, S. (2013). Psychopathic personality and utilitarian moral judgment in college students. *Journal of Criminal Justice, 41*(5), 342–349. https://doi.org/10.1016/j.jcrimjus.2013.06.012

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology*, *113*(3), 343. https://psycnet.apa.org/doi/10.1037/pspa0000086

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS One, 8*(4), e60418. https://doi.org/10.1371/journal.pone.0060418

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology, 45*(3), 581-584. https://doi.org/10.1016/j.jesp.2009.01.003

Greene, J. D. (2023). Dual-process moral judgment beyond fast and slow. *Behavioral and Brain Sciences*, *46*, e123. https://doi.org/10.1017/S0140525X22003193

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. https://psycnet.apa.org/doi/10.1037/0033-295X.108.4.814

Haidt, J. (2003). The Moral Emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (eds.). *Handbook of Affective Sciences.* Oxford: Oxford University Press.

Hare, R. D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior*, *23*(1), 25-54. https://doi.org/10.1177%2F0093854896023001004

Hare, R. D., & Neumann, C. S. (2009). Psychopathy: Assessment and forensic implications. *The Canadian Journal of Psychiatry*, *54*(12), 791-802. https://doi.org/10.1177%2F070674370905401202

Herpertz, S. C., Werth, U., Lukas, G., Qunaibi, M., Schuerkens, A., Kunert, H. J., ... & Sass, H. (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Archives of General Psychiatry*, *58*(8), 737-745. https://doi.org/10.1001/archpsyc.58.8.737

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*, 307–316. https://psycnet.apa.org/doi/10.1037/h0027580

Hojat, M., Gonnella, J. S., Nasca, T. J., Mangione, S., Veloksi, J. J., & Magee, M. (2002). The Jefferson Scale of Physician Empathy: further psychometric data and differences by gender and specialty at item level. *Academic Medicine*, 77(10), S58-S60.

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas? *PloS One*, *11*(7), e0154780. https://doi.org/10.1371/journal.pone.0154780

Jansen, S. J. T. (2011). The Multi-attribute Utility Method. In *The Measurement and Analysis of Housing Preference and Choice*, edited by S. J. T. Jansen, H. C. C. H. Coolen, R. W. Goetgeluk, 101–125. Dordrecht: Springer.

Jansen, S., Boumeester, H., Coolen, H., Goetgeluk, R., & Molin, E. (2009). The impact of including images in a conjoint measurement task: Results of two small-scale studies. *Housing and the Built Environment*, *24*(3), 271–297.

Johnson, R. M. (1987). *Adaptive Conjoint Analysis*. Sawtooth Software Conference Proceedings, 253–65.

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the basic empathy scale. *Journal of Adolescence, 29*, 589-611. https://doi.org/10.1016/j.adolescence.2005.08.010

Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131–164. https://doi.org/10.1037/rev0000093

Kahane, G., Everett, J. A., Earp, B. D., Farias, M., & Savulescu, J. (2015). Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition*, *134*, 193–209. https://doi.org/10.1016/j.cognition.2014.10.005

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402. https://doi.org/10.1093/scan/nsr005

Kahneman, D. (2013). *Thinking, fast and slow*. Farrar, Straus, and Giroux.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*(1), 160-173. https://www.jstor.org/stable/24966506

Kant, I. (1785). *Grundlegung zur Metaphysik der Sitten*. Riga: Johann Friedrich Hartknoch.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Koenigs, M., Kruepke,M., Zeier, J., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience, 7*(6), 708–714. https://doi.org/10.1093/scan/nsr048

Li, Y., & Liu, A.-S. (2017). Moral judgment and its relationship with theory of mind in patients with autism spectrum disorder. *Chinese Mental Health Journal, 31*(5), 366–370.

Lotto, L., Manfrinati, A., & Sarlo, M. (2014). A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. *Journal of Behavioral Decision Making, 27*, 57–65. https://doi.org/10.1002/bdm.1782.

May, J. (2018). The Limits of Emotion in Moral. In K. Jones & F. Schroeter (Eds.), *The many moral rationalisms* (pp. 286-306). Oxford University Press.

Mill, J. S. (1979).*Utilitarianism*. Indianapolis, IN: Hackett. (Original work published 1861)

Miller, R., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass, 7*(10), 707–718. https://doi.org/10.1111/spc3.12066

Nasello, J. A., & Triffaux, J. M. (2023a). A stenography of empathy: Toward a consensual model of the empathic process. *L'Encephale*, *49*(4), 399-407. https://doi.org/10.1016/j.encep.2022.12.002

Nasello, J. A., & Triffaux, J. M. (2023b). The role of empathy in trolley problems and variants: A systematic review and meta-analysis. *British Journal of Social Psychology, 00*, 1-29. https://doi.org/10.1111/bjso.12654

Nasello, J. A., Blavier, A., & Triffaux, J. M. (2021b). French adaptation of the Five-Factor Borderline Inventory-Short Form. *Current Psychology*, 1-12. https://doi.org/10.1007/s12144-021-01878-2

Nasello, J. A., Dardenne, B., Blavier, A., & Triffaux, J. M. (2021a). Does empathy predict decision-making in everyday trolley-like problems?. *Current Psychology*, 1-14. https://doi.org/10.1007/s12144-021-01566-1

Nasello, J. A., Dardenne, B., Hansenne, M., Blavier, A., & Triffaux, J. M. (2023a). Moral decision-making in trolley problems and variants: how do participants' perspectives, borderline personality traits, and empathy predict choices?. *The Journal of Psychology, 157*(5), 318-338. https://doi.org/10.1080/00223980.2023.2206604

Nasello, J., Triffaux, J. M., & Hansenne, M. (2023b). Individual differences and personality traits across situations. *Current Issues in Personality Psychology*. https://doi.org/10.5114/cipp/159942

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem". *Emotion, 12*(2), 364–370. https:// doi.org/10.1037/a0025561

Oudman, E., van Stigt Thans, S., Montoya, E. R., & Postma, A. (2021). Moral reasoning, moral decision-making, and empathy in Korsakoff's syndrome. *Journal of Neuropsychology*, *15*(3), 462-476. https://doi.org/10.1111/jnp.12233

Patil, I. (2015). Trait psychopathy and utilitarian moral judgment: The mediating role of action aversion. *Journal of Cognitive Psychology*, *27*(3), 349–366. https://doi.org/10.1080/20445911.2015.1004334.

Patil, I., & Silani, G. (2014). Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology, 5*, 501. https://doi.org/10.3389/fpsyg.2014.00501

Patil, I., Larsen, E. M., Kichic, R., & Gleichgerrcht, E. (2021). Moral cognition about harm in anxiety disorders: The importance of experienced emotion. *Psychological Reports, 124*(6), 2501-2523. https://doi.org/10.1177/0033294120964134

Patil, I., Melsbach, J., Hennig-Fast, K., & Silani, G. (2016). Divergent roles of autistic and alexithymic traits in utilitarian moral judgments in adults with autism. *Scientific Reports*, *6*(1), 1-15. https://doi.org/10.1038/srep23637

Peckham, A. D., Jones, P., Snorrason, I., Wessman, I., Beard, C., & Björgvinsson, T. (2020). Age-related differences in borderline personality disorder symptom networks in a transdiagnostic sample. *Journal of Affective Disorders*, *274*, 508-514. https://doi.org/10.1016/j.jad.2020.05.111

Pletti, C., Lotto, L., Buodo, G., & Sarlo, M. (2017). It's immoral, but I'd do it! Psychopathy traits affect decision-making in sacrificial dilemmas and in everyday moral situations. British Journal of Psychology, 108(2), 351-368. https://doi.org/10.1111/bjop.12205

Powell, L. J. (2022). Adopted utility calculus: Origins of a concept of social affiliation. *Perspectives on Psychological Science*, *17*(5), 1215-1233.

Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Science*, *25*(1), 1-20. https://doi.org/10.1017/S0140525X02000018

Richetin, J., Preti, E., Costantini, G., & De Panfilis, C. (2017). The centrality of affective instability and identity in Borderline Personality Disorder: Evidence from network analysis. *PloS One*, *12*(10), e0186695. https://doi.org/10.1371/journal.pone.0186695

Salgado, R. M., Pedrosa, R., & Bastos-Leite, A. J. (2020). Dysfunction of empathy and related processes in borderline personality disorder: a systematic review. *Harvard review Review of Psychiatry*, *28*(4), 238. https://doi.org/10.1097/HRP.0000000000000260

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818-831. https://doi.org/10.1177/00131644950550050

Sidgwick, H., 1874. *The methods of ethics*. London: Macmillan.

Smillie, L. D., Katic, M., & Laham, S. M. (2020). Personality and moral judgment: Curious consequentialists and polite deontologists. *Journal of Personality*, *89*(3), 549-564. https://doi.org/10.1111/jopy.12598

Southward, M. W., & Cheavens, J. S. (2018). Identifying core deficits in a dimensional model of borderline personality disorder features: A network analysis. *Clinical Psychological Science*, *6*(5), 685-703. https://doi.org/10.1177%2F2167702618769560

Struett, D. (2023). Husband of Elk Grove mom killed in Florida parasailing crash warns of shady tourist companies. *Chicago Sun*Times*. Retrieved from : https://chicago.suntimes.com/2023/6/8/23754092/parasailing-crash-elk-grove-village-florida-keys-srinivasrao-supraja-alaparthi#:~:text=The%2033%2Dyear%2Dold%20mother,claims%20caused%20his%20wife's%20death

Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre-and post-decision processes. *Acta Psychologica, 80*(1-3), 143-168. https://doi.org/10.1016/0001-6918(92)90044-E

Takamatsu, R. (2019). Personality correlates and utilitarian judgments in the everyday context: Psychopathic traits and differential effects of empathy, social dominance orientation, and dehumanization beliefs. *Personality and Individual Differences, 146*, 1–8. https://doi.org/10. 1016/j.paid.2019.03.029

Takamatsu, R., & Takai, J. (2019). With or without empathy: Primary psychopathy and difficulty in identifying feelings predict utilitarian judgment in sacrificial dilemmas. *Ethics & Behavior*, *29*(1), 71–85. https://doi.org/10.1080/10508422.2017.1367684

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59*, 204–217. https://www.jstor.org/stable/27902416

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

Vyas, K., Jameel, L., Bellesi, G., Crawford, S., & Channon, S. (2017). Derailing the trolley: Everyday utilitarian judgments in groups high versus low in psychopathic traits or autistic traits. *Psychiatry Research, 250*, 84–91. https://doi.org/10.1016/j.psychres.2017.01. 054.

Yu, C. L., & Chou, T. L. (2018). A dual route model of empathy: A neurobiological prospective. *Frontiers in Psychology, 9*, 2212. https://doi.org/10.3389/fpsyg.2018.02212

Zhan, Y., Xiao, X., Li, J., Liu, L., Chen, J., Fan, W., & Zhong, Y. (2018). Interpersonal relationship modulates the behavioral and neural responses during moral decision-making. *Neuroscience Letters*, *672*, 15-21. https://doi.org/10.1016/j.neulet.2018.02.039

Zhang, X., Wu, Z., Li, S., Lai, J., Han, M., Chen, X., Liu, C., & Ding, D. (2020). Why people with high alexithymia make more utilitarian judgments: The role of empathic concern and deontological inclinations. *Experimental Psychology*, *67*(1), 23–30. https://doi.org/10.1027/1618-3169/a000474

Section 4 : Annexes

Annexe A : Etude 1, everyday trolley-like problem, le dilemme du tramway (version université)

Scenario 1, *impersonal* scenario (*individual- versus group*-oriented choices)

Perspective : The tramway driver

Vous êtes une conductrice de tramway menant à une Université. Ce tramway est complet (nombre maximal de personnes assises et debout) et comporte un ensemble d'étudiants se rendant à différents cours à présence obligatoire de cette Université. L'Université étant excentrée, le trajet est assez long (1h30). De ce fait, aucun arrêt n'est envisagé et l'horaire prévoit une navette toutes les heures trente. Ainsi, chaque navette est paramétrée pour permettre aux usagers d'arriver à l'heure en cours.

L'Université ayant une politique stricte sur la présence des étudiants aux cours, peu de retards et absences sont tolérés. D'une part, au-delà d'une certaine limite de retard, certains professeurs peuvent refuser que les étudiants assistent au cours ; d'autre part, les étudiants qui n'assistent pas au cours pourraient se voir refuser l'accès à l'examen du cours manqué en première session.

Après plusieurs kilomètres de route, vous recevez un appel de votre centrale vous avertissant qu'un(e) passager(ère) (d'environ votre âge) attend sur un quai au milieu du trajet avec son sac de cours bien rempli. Exceptionnellement, vous pouvez choisir de vous arrêter ou non pour charger cette passagère.

Si vous choisissez de vous arrêter, on estime un retard de 15 minutes, plusieurs passagers seront donc en retard à leurs cours (à la limite de la fermeture définitive de l'accès à l'auditoire). Pour les étudiants dont les locaux sont plus éloignés et ceux dont les professeurs sont les plus à cheval sur les horaires, certains pourraient peut-être manquer leur cours (et donc potentiellement manquer l'accès à leur examen en première session). De plus, les autres passagers bénéficieront d'un confort moindre car ils devront se serrer davantage pour le reste du trajet.

Toutefois, cela permettra à la nouvelle passagère d'arriver à son cours de manière certaine, sans risque de pénalité quelconque.

Si vous choisissez de ne pas vous arrêter, la passagère manquera de manière certaine son cours à présence obligatoire mais les autres voyageurs arriveront à temps aux leurs.

Pour cette passagère, comme pour les passagers déjà présents, il est complexe de justifier un retard auprès d'un professeur et beaucoup d'entre eux préfèrent « montrer l'exemple » en pénalisant les étudiants qui ne sont pas à l'heure. Enfin, sachant que cette passagère présente une assistance au cours obligatoire qui est identique aux autres passagers du tramway, que décider ?

N.B. : d'autres moyens de transports (e.g. taxi, bus, car ou autre) ne sont pas envisageables car, comme mentionné au départ, il s'agit d'une université excentrée où peu de trafic afflue. La passagère du quai dispose d'une excuse que vous qualifiez de valable pour ne pas avoir démarré au point de départ prévu.

Que décidez-vous ?

- ○ Arrêter le tramway
- ○ Ne pas arrêter le tramway

Option alternative :

En vous arrêtant, vous pouvez choisir également d'imposer ou non des conditions spécifiques à cet/cette étudiant(e) ou décider de maintenir votre choix de ne pas arrêter la navette. Les deux conditions sont : 1) imposition d'un tarif supplémentaire et 2) la passager(ère) doit laisser son sac de cours dans une consigne du quai pour éviter d'encombrer davantage les passagers déjà présents (coûts supplémentaires de consigne et récupération complexe du sac après les cours). Le strict nécessaire sera donc pris. Dans cette condition, la voyageuse est avertie préalablement par la centrale afin d'occasionner moins de retard.

Choisissez une des options suivantes :

- J'arrête le tramway pour charger cet/cette étudiant(e) mais je lui impose les conditions différenciées.
- J'arrête le tramway pour charger cet/cette étudiant(e) mais je ne lui impose pas les conditions différenciées.
- Je n'arrête pas le tramway pour charger cet/cette étudiant(e).

Scenario 2, personal scenario (*self- versus group*-oriented choices)

Perspective : The student in a mid-trip platform

Vous êtes un(e) passager(ère) qui attend sur un quai, au milieu d'un trajet de tramway menant à votre Université. Ce tramway est complet (nombre maximal de personnes assises et debout) et comporte un ensemble d'étudiants se rendant à différents cours à présence obligatoire de cette Université. L'Université étant excentrée, le trajet est assez long (1h30). De ce fait, aucun arrêt n'est envisagé et l'horaire prévoit une navette toutes les heures trente. Ainsi, chaque navette est paramétrée pour permettre aux usagers d'arriver à l'heure en cours.

L'Université ayant une politique stricte sur la présence des étudiants aux cours, peu de retards et absences sont tolérés. D'une part, au-delà d'une certaine limite de retard, certains professeurs peuvent refuser que les étudiants assistent au cours ; d'autre part, les étudiants qui n'assistent pas au cours pourraient se voir refuser l'accès à l'examen du cours manqué en première session.

Vous vous trouvez donc au milieu du trajet sur un quai. Vous avez pris la décision de contacter la centrale afin de leur expliquer la raison valable pour laquelle vous vous êtes retrouvé(e) dans cette situation (manquement du départ du tramway). Exceptionnellement, la centrale vous informe que vous pouvez choisir de faire arrêter ou non le tramway afin de pouvoir y monter.

Si vous choisissez de faire arrêter le tramway, on estime un retard de 15 minutes, plusieurs passagers seront donc en retard à leurs cours (à la limite de la fermeture définitive de l'accès à l'auditoire). Pour les étudiants dont les locaux sont plus éloignés et ceux dont les professeurs sont les plus à cheval sur les horaires, certains pourraient peut-être manquer leur cours (et donc potentiellement manquer l'accès à leur examen en première session). De plus, les autres passagers bénéficieront d'un confort moindre car ils devront se serrer davantage pour le reste du trajet. Toutefois, cela vous permettra d'arriver à votre cours de manière certaine, sans risque de pénalité quelconque.

Si vous choisissez de ne pas faire arrêter le tramway, vous manquerez de manière certaine votre cours à présence obligatoire mais les autres voyageurs arriveront à temps aux leurs.

Pour vous, comme pour les passagers déjà présents, il est complexe de justifier un retard auprès d'un professeur et beaucoup d'entre eux préfèrent « montrer l'exemple » en pénalisant les étudiants qui ne sont pas à l'heure. Enfin, sachant que vous avez une présence obligatoire au cours identique aux autres passagers du tramway, que décider ?

N.B. : d'autres moyens de transports (e.g. taxi, bus, car ou autre) ne sont pas envisageables car, comme mentionné au départ, il s'agit d'une université excentrée où peu de trafic afflue. Vous disposez d'une excuse que vous qualifiez de valable pour ne pas avoir démarré au point de départ prévu.

Que décidez-vous ?

- ○ Arrêter le tramway
- ○ Ne pas arrêter le tramway

Option alternative :

En vous arrêtant, vous pouvez choisir également de vous imposer ou non des conditions spécifiques ou décider de maintenir votre choix de ne pas faire arrêter la navette. Les deux conditions sont : 1) imposition d'un tarif supplémentaire et 2) déposer son sac de cours dans une consigne du quai pour éviter d'encombrer davantage les passagers déjà présents (coûts supplémentaires de consigne et récupération complexe du sac après les cours). Le strict nécessaire sera donc pris. Dans cette condition, vous êtes averti(e) préalablement par la centrale afin d'occasionner moins de retard.

<u>Choisissez une des options suivantes :</u>

- ○ Je fais arrêter le tramway pour être chargé(e) mais je m'impose les conditions différenciées.
- ○ Je fais arrêter le tramway pour être chargé(e) mais je ne m'impose pas les conditions différenciées.
- ○ Je ne fais pas arrêter le tramway pour être chargé(e).

Annexe B : Basic Empathy Scale (Jolliffe & Farrington, 2006 ; version française : D'Ambrosio et al., 2009)

1 = Fortement en désaccord	2 = En désaccord	3 = Neutre	4 = En accord	5 = Fortement en accord

#	Items	Importance				
1	Les émotions de mes ami(e)s ne m'affectent pas beaucoup.	1	2	3	4	5
2	Après avoir été avec un(e) ami(e) qui est triste, je me sens généralement triste.	1	2	3	4	5
3	Je peux comprendre le bonheur que ressent un(e) ami(e) quand quelque chose de bien lui arrive.	1	2	3	4	5
4	Je peux être effrayé(e) par un bon film d'horreur.	1	2	3	4	5
5	Je m'imprègne facilement des émotions des autres.	1	2	3	4	5
6	J'ai du mal à savoir quand mes ami(e)s ont peur.	1	2	3	4	5
7	Je ne suis pas triste quand je vois une personne pleurer.	1	2	3	4	5
8	Les sentiments des autres ne me tracassent pas du tout.	1	2	3	4	5
9	Quand quelqu'un a la sensation "d'être au plus bas", je peux comprendre ce qu'il ressent.	1	2	3	4	5
10	Généralement, je peux me rendre compte quand mes ami(e)s sont effrayé(e)s.	1	2	3	4	5
11	Souvent je me sens triste lorsque je regarde des choses ou des films tristes.	1	2	3	4	5
12	Souvent je peux comprendre comment les gens se sentent avant même qu'ils me l'aient dit.	1	2	3	4	5
13	Voir une personne se mettre en colère n'a aucun effet sur moi.	1	2	3	4	5
14	Habituellement, je sais quand les gens sont joyeux.	1	2	3	4	5
15	J'ai tendance à me sentir apeuré(e) quand je suis avec des ami(e)s qui sont effrayé(e)s.	1	2	3	4	5
16	Généralement, je me rends vite compte quand un(e) ami(e) est en colère.	1	2	3	4	5
17	Je suis souvent envahi(e) par les sentiments de mes ami(e)s.	1	2	3	4	5
18	Je ne ressens rien face à la tristesse de mes ami(e)s.	1	2	3	4	5
19	Je ne suis généralement pas attentif(ve) aux sentiments de mes ami(e)s.	1	2	3	4	5
20	J'ai du mal à comprendre quand mes ami(e)s sont heureux(ses).	1	2	3	4	5

Annexe C : Five-Factor Borderline Inventory, Short Form

Version française : Nasello, Blavier, & Triffaux, 2021.

Nasello, J., Blavier, A., & Triffaux, J-M. (2021). French adaptation of the Five-Factor Borderline Inventory-Short Form. *Current Psychology*, 1-12. https://doi.org/10.1007/s12144-021-01878-2

Ce questionnaire comprend 48 énoncés qui évaluent la manière dont vous agissez, pensez et ressentez. Chaque énoncé est à évaluer sur une échelle de Likert allant de 1 à 5 : **1** correspondant à "**Totalement en désaccord**" ; **2** correspondant à "**Plutôt en désaccord**" ; **3** correspondant à "**Ni en accord, ni en désaccord**" ; **4** correspondant à "**Plutôt en accord**" ; **5** correspondant à "**Totalement en accord**".

Pour chaque item, il faut donc attribuer un degré d'accord ou de désaccord qui traduit à quel point cet item vous correspond. En répondant, **représentez-vous votre manière d'être de façon générale**, sans vous référer à un événement exclusif ou spécifique. Enfin, rappelez-vous qu'il n'y a pas de bonnes ou mauvaises réponses.

Items	**Evaluation**				
1. *J'ai tendance à être assez anxieux(se).*	1	2	3	4	5
2. *J'ai déjà fait quelques crises de colère.*	1	2	3	4	5
3. *Je me sens parfois inutile et sans valeur.*	1	2	3	4	5
4. *Je peux être tellement différent en fonction des personnes, comme si je n'étais pas la même personne.*	1	2	3	4	5
5. *J'ai fréquemment des envies irrépressibles de faire des choses qui me causent des ennuis.*	1	2	3	4	5
6. *Mes émotions peuvent devenir incontrôlables.*	1	2	3	4	5
7. *Me faire du mal est l'une des seules façons qui me permette de tolérer mes émotions.*	1	2	3	4	5
8. *J'ai déjà ressenti que des choses étaient irréelles et que j'étais détaché(e) de la vie (ex : impression d'être déconnecté(e) de mon corps).*	1	2	3	4	5
9. *Je me méfie souvent des autres.*	1	2	3	4	5
10. *Je fais parfois des choses que je ne devrais pas pour obtenir des autres qu'ils fassent ce dont j'ai envie ou besoin.*	1	2	3	4	5
11. *J'ai tendance à entrer fréquemment en conflit avec les*	1	2	3	4	5

autres.					
12. *Je m'attire des problèmes car je n'anticipe pas les conséquences de mes actions.*	1	2	3	4	5
13. *Je m'inquiète énormément.*	1	2	3	4	5
14. *Ma colère me semble souvent incontrôlable.*	1	2	3	4	5
15. *Il m'est arrivé de réfléchir à des moyens de me suicider.*	1	2	3	4	5
16. *Je peux être si différent(e) en fonction des personnes que je me demande qui je suis.*	1	2	3	4	5
17. *Parfois, je me laisse emporter par mes envies irrépressibles.*	1	2	3	4	5
18. *J'ai le sentiment de ne pas avoir beaucoup de contrôle sur ce que je ressens.*	1	2	3	4	5
19. *J'ai déjà menacé de me suicider.*	1	2	3	4	5
20. *J'ai parfois l'impression de ne plus être connecté(e) à mon corps.*	1	2	3	4	5
21. *C'est vraiment difficile pour moi de faire confiance aux gens.*	1	2	3	4	5
22. *Des personnes m'ont déjà qualifié de manipulateur(trice).*	1	2	3	4	5
23. *Je menacerais des gens pour les obliger à faire ce que je veux.*	1	2	3	4	5
24. *J'ai tendance à agir rapidement sans me poser de questions.*	1	2	3	4	5
25. *J'ai très peur que les gens me quittent ou me délaissent.*	1	2	3	4	5
26. *Ma colère prend parfois le dessus sur moi.*	1	2	3	4	5
27. *Je me sens souvent triste.*	1	2	3	4	5
28. *J'ai l'impression de n'être proche de personne.*	1	2	3	4	5
29. *Quand je suis contrarié(e) ou énervé(e), je fais souvent des choses qui me posent plus tard des problèmes.*	1	2	3	4	5
30. *Mon humeur bascule rapidement d'un sentiment à l'autre.*	1	2	3	4	5
31. *Des échecs, même mineurs, peuvent me causer un grand désarroi dans ma vie.*	1	2	3	4	5
32. *J'ai parfois l'impression de ne pas être réel.*	1	2	3	4	5
33. *Les gens ne me sont pas aussi fidèles et loyaux que je le voudrais.*	1	2	3	4	5
34. *Je suis connu pour dissimuler ou embellir la vérité afin d'arriver à mes fins.*	1	2	3	4	5
35. *Je rentre souvent en conflit avec les personnes qui me sont*	1	2	3	4	5

proches.					
36. *Les autres disent de moi que je ne réfléchis pas avant d'agir.*	1	2	3	4	5
37. *Je m'inquiète énormément pour les choses qui sont hors de mon contrôle.*	1	2	3	4	5
38. *Ma colère m'a parfois causé des ennuis.*	1	2	3	4	5
39. *Je pense au suicide depuis mon adolescence.*	1	2	3	4	5
40. *Je me sens souvent comme exclu(e) ou rejeté(e) par les autres.*	1	2	3	4	5
41. *J'ai déjà fait beaucoup de choses impulsivement que j'ai regrettées ensuite.*	1	2	3	4	5
42. *J'ai du mal à contrôler mes humeurs.*	1	2	3	4	5
43. *Je ne pense pas pouvoir continuer à vivre de cette façon.*	1	2	3	4	5
44. *J'ai parfois l'impression que rien n'est réel.*	1	2	3	4	5
45. *Je ne parviens pas à faire confiance à certain(e)s de mes ami(e)s les plus proches.*	1	2	3	4	5
46. *Parfois, il faut être manipulateur(trice) et malhonnête pour obtenir ce dont on a besoin.*	1	2	3	4	5
47. *Je suis facile à vivre.*	1	2	3	4	5
48. *J'ai déjà fait de très mauvaises choses sous le coup de l'impulsivité.*	1	2	3	4	5

Cotation :

Score total :

- Etape 1 : inverser l'item 47 (6 – X). X = degré d'accord ou de désaccord sélectionné par le.la patient.e/participant.e (e.g., *Je suis facile à vivre*. Réponse du participant : 4 ; faire 6-4= 2) ;
- Etape 2 : addition de tous les items ; plus le score est élevé, plus les traits de personnalité sont marqués ;
- Etape 3 : comparaison du score total avec score seuil (**162**) → si le.la participant.e obtient un score **supérieur ou égal à 162**, cela reflète un individu présentant des traits borderlines saillants.
    - **Attention : ce score est un indicateur. En aucun cas, il ne permet de réaliser à lui seul un diagnostic psychopathologique. L'objectif étant d'étayer une hypothèse clinique et/ou de cibler les difficultés du.de la patient.e, parmi les différentes sous-dimensions proposées par l'échelle.**
- Etape 4 : calculer les scores aux sous-dimensions : additionner les items de chaque dimension (Cf. ci-dessous) ; plus le score est élevé, plus la dimension est importante.

Sous-dimensions :

Douze sous-dimensions sont calculées (additionner les quatre items associés à chaque sous-dimension ; plus les scores sont élevés, plus les difficultés sont marquées)

- **Incertitude Anxieuse** (*Anxious Uncertainty*) (N1) : 1 + 13 + 25 + 37
- **Colère Dysfonctionnelle** (*Dysregulated Anger*) (N2) : 2 + 14 + 26 + 38
- **Profond Découragement** (*Despondence*) (N3) : 3 + 15 + 27 + 39
- **Perturbation subjective de l'identité** (Self-Disturbance) (N4) : 4 + 16 + 28 + 40
- **Perturbation Comportementale** (*Behavioral Dysregulation*) (N5) : 5 + 17 + 29 + 41
- **Perturbation Affective** (*Affective Dysregulation*) (N6a) : 6 + 18 + 30 + 42
- **Fragilité** (*Fragility*) (N6b) : 7 + 19 + 31 + 43
- **Tendances Dissociatives** (*Dissociative Tendencies*) (O1) : 8 + 20 + 32 + 44
- **Méfiance** (*Distrustfulness*) (A1) : 9 + 21 + 33 + 45
- **Manipulation** (*Manipulativeness*) (A2) : 10 + 22 + 34 + 46 + 23
- **Oppositionalité** (*Oppositional*) (A4) : 11 + 35 + 47(R)
- **Précipitation** (*Rashness*) (C6) : 12 + 24 + 36 + 48

Annexe D : traduction française de l'échelle « Five-Factor Borderline Inventory, Short-form »

Current Psychology
https://doi.org/10.1007/s12144-021-01878-2

French adaptation of the Five-Factor Borderline Inventory-Short Form

Julian A. Nasello[1,2] · Adélaïde Blavier[2] · Jean-Marc Triffaux[1,3]

Accepted: 18 May 2021

Abstract

The purpose of this study was to adapt the short-form version of the Five-Factor Borderline Inventory (FFBI-SF) in French, a 12-domain questionnaire of 48 items assessing Borderline Personality Disorders (BPD). Participants from the general population voluntarily took part in our study and were separated into two samples: a student ($n = 335$) and a worker sample ($n = 162$). They completed four questionnaires randomly distributed: a demographic questionnaire, the FFBI-SF, the Big Five Inventory, the McLean Screening Instrument for Borderline Personality Disorder, and the Borderline Symptom List, Short Form. Our results showed acceptable model fit indices for a 12-factor model and acceptable to excellent reliability indices (overall $\alpha = .95$ and average α of all domains is .78). Strong correlations were found between FFBI-SF and other BPD measures, suggesting good convergent validity. Moreover, compared with other BPD questionnaires, the incremental validity of the FFBI-SF was between 13 and 23%, showing a substantial additional part of variance explained by the FFBI-SF. Our findings provide evidence for a valid and reliable French adaptation of the FFBI-SF. This study fits with a growing need for clinicians and researchers to benefit from multifacet, accessible, and quickly completed tools that assess subclinical BPD.

Keywords Borderline personality disorders · Five-Factor Borderline Inventory · Five-Factor Model · Psychiatry · Psychopathology

For decades, Borderline Personality Disorder (BPD) has raised research interests. BPD is a severe psychiatric disorder (Bender et al., 2001; Zanarini et al., 2012), and its lifetime prevalence is estimated to be between 1 and 3% in the general population (Leichsenring et al., 2011). In psychiatric settings, this prevalence raised at 10–11% for outpatients (Gunderson, 2011) and 33% for inpatients (Ha et al., 2014). According to the Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition (DSM-V; American Psychiatric Association, 2013, p.663), BPD is characterized by a pervasive pattern of instability of interpersonal relationships, self-image, and affects, and marked impulsivity. Nine criteria are listed, requiring at least five to receive the BPD diagnosis. However, the current categorical approach of the DSM-V becomes less popular, and research evidence favors the validity of the dimensional over the categorical approach[1] (Widiger & Trull, 2007). It is especially true for personality disorders in general (see a complete comparison between categorical and dimensional approaches in Alfonso et al., 2021) and BPD. Indeed, several authors consider BPD as a dimensional construct (Carvalho et al., 2018; Hopwood et al., 2018; Kotov et al., 2017), and several measures have been created following this dimensional perspective. This study focused on the Five-Factor Borderline Inventory, short-form (FFBI-SF; DeShong et al., 2016), and aimed to adapt this scale in French.

The FFBI was developed based on the Five-Factor Model (FFM; Costa Jr & McCrae, 1992; Costa Jr & McCrae, 1995). This model includes five big domains of personality (i.e., extraversion, agreeableness, openness to experience, conscientiousness, and neuroticism), and each domain is composed of

✉ Julian A. Nasello
julian.nasello@gmail.com

[1] Psychosomatic Medicine and Group Psychotherapy, Psychiatric Day Hospital « La Clé », Boulevard de la Constitution 153, 4020 Liège, Belgium

[2] Department of Clinical Psychology, University of Liège, Liège, Belgium

[3] Department of Psychiatry, Medicine, University of Liège, CHU of Liège, Liège, Belgium

[1] It is important to note DSM-V efforts to adopt a transition from a categorical to a dimensional approach. For instance, an alternative hybrid model has been proposed for personality disorders (see DSM-V, Section III). Besides, DSM-V mentioned that, in coming years, dimensional approaches will probably supplement or supersede current categorical approaches.

Published online: 02 June 2021

six facets. The FFM has demonstrated relevant predictions of all ten personality disorders, and these personality disorders can be understood as extreme and maladaptive expressions of the big five domains (Samuel & Widiger, 2008; for BPD: Miller et al., 2012; Trull & Brown, 2013; Helle et al., 2018). Mullins-Sweatt et al. (2012) created the FFBI as follow: seven subscales were based on the neuroticism facets (i.e., Anxious Uncertainty, Dysregulated Anger, Despondence, Self-Disturbance, Behavioral Dysregulation, Affective Dysregulation, and Fragility; one subscale was based on the openness (i.e., Dissociative Tendencies); three subscales were based on the facets of agreeableness (i.e., Distrustfulness, Manipulativeness, and Oppositional); and, lastly, one subscale was based on the conscientiousness facet (i.e., Rashness). DeShong et al. (2016) introduced a 48-item short form of the FFBI (Mullins-Sweatt et al., 2012). This scale presents significant benefits (see details in DeShong et al., 2016). First, there is a substantial reduction in time for completion. Participants complete the short-form questionnaire in 5 or 10 min, representing a significant interest for clinicians or researchers. Second, the short form is reliable (the average coefficient alpha of all domains is .78). Third, the short form presents strong convergent and discriminant validity. Four, the multidimensional approach of the FFBI (long and short forms) can easily build a comprehensive profile of the most central and peripheral components of BPD. Lastly, the FFBI model and general personality are linked (as mentioned, the FFBI was based on the Five-Factor Model and understood BPD as maladaptive personality traits). Therefore, it can help understand the etiology, development, and treatment of BPD (DeShong et al., 2016).

The present study aimed to adapt the FFBI-SF in French and explore its psychometric properties to determine whether this adaptation is reliable and valid.

Method

Participants and Procedure

Participants were collected through an online survey on a voluntary basis. They indicated their informed consent by clicking on the "*next page*" option after reading complete information about the study. Individuals who did not want to participate or give their informed consent had to close their browser and were automatically exited from the study. At the end of their participation, they received a printable debriefing. The participants were not offered any course credit or compensation for responding. Two non-clinical samples of participants were used in this anonymous study. A sample composed of students (S1: $N = 335$; $M_{age} = 21.9$; $SD_{age} = 3.12$) and another one composed of workers (S2: $N = 162$; $M_{age} = 31.7$; $SD_{age} = 7.5$). Most participants from Sample 1 (composed of 127 men and 208 women) were not financially independent (87.5%), described their parents' socioeconomic status as "on average" (51.3%) or "higher than average" (34%), and. 63.3% were bachelor students, and 31.9% were Master students. Sample 2 was composed of 72 men and 90 women (56%). Most participants were full-time workers (72.2%) and described their socioeconomic status as "on average" (44.4%) or "higher than average" (29.6%). 56.8% had a Master's, and 22.8% a Bachelor's degree. For both samples, we asked participants to report neurological and psychiatric histories (e.g., "*Have you experienced or are you experiencing neurological/psychiatric disorders (diagnosed by a psychiatrist/psychologist/neurologist)?*"). Then, if they answered *Yes*, participants had to describe on a 5-point Likert scale the current importance of difficulties they are experiencing, from "1″ (*No current difficulty*) to "5″ (*Severe current difficulties*). Participants who selected "≥ 4″ were excluded (*Important* or *Severe current difficulties*). Eight participants were excluded from analyses in Sample 1 and three in Sample 2.

The ethical committee of the Department of Psychology of the University of Liège (Belgium) approved the study, reference n°: 1920-92.

Measures

Participants from both samples received the same questionnaires. They had to fulfill a basic demographic questionnaire (information includes age, gender, relationship status, current occupation, year in school, department, neurological and psychiatric histories.), the Five-Factor Borderline Inventory, short-form (FFBI-SF; DeShong et al., 2016) translated in French, the Big Five Inventory (BFI; John & Srivastava, 1999; French version: Plaisant et al., 2010); the McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD; Zanarini et al., 2003; French version: Mirkovic et al., 2020); and the Borderline Symptom List, short from (BSL-23; Bohus et al., 2001; Bohus et al., 2009; French version: Nicastro et al., 2016).

Five-Factor Borderline Inventory, Short Form The FFBI-SF (DeShong et al., 2016) is a 48-item self-rated measure. Each item is answered on a 5-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). This short version was constructed based on a longer version of 120 items (FFBI; Mullins-Sweatt et al., 2012). Like its more extended version, the scale consists of assessing 12 subscales. Each subscale is composed of 4 items[2]: Anxious Uncertainty (items 1, 13, 25, 37; e.g., "*I tend to be quite anxious*"); Dysregulated Anger (items 2, 14, 26, 38; e.g., "*I have already had some angry outbursts*"); Despondence (items 3, 15, 27, 39; e.g., "*I sometimes feel*

[2] DeShong et al. (2016) performed IRT analyses and found that the four-item scale was the best option in comparison with the three- and five-scales.

Table 1 Internal consistency and descriptive statistics

Scale	Subscales	α		Mean		SD		Min		Max	
		S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
FFBI-SF	Total	.95	.95	114	106	31.3	30	56	49	212	174
	Anx. Unc. (N1)	.80	.81	13.7	12.6	4.07	4.2	4	4	20	20
	Women			14.7	13.7	4.07	3.99	5	4	20	20
	Men			11.9	11.1	3.62	4.18	4	4	20	19
	Dys. Ang. (N2)	.88	.87	9.84	9.31	4.49	4.1	4	4	20	20
	Desp. (N3)	.80	.85	10.3	9.11	4.07	4.12	4	4	20	20
	Self-Dist. (N4)	.76	.76	10.3	9.49	3.99	3.73	4	4	20	18
	Behav. Dys. (N5)	.77	.80	8.7	8.23	3.63	3.51	4	4	19	18
	Aff. Dys. (N6a)	.86	.85	11	9.87	4.29	3.99	4	4	20	19
	Women			11.6	10.5	4.16	4.03	4	4	20	19
	Men			10	9.13	4.34	3.84	4	4	20	18
	Frag. (N6b)	.67	.66	8.07	7.28	3.2	2.92	4	4	19	17
	Diss. Tend. (O1)	.80	.84	7.69	6.84	4.26	3.63	4	4	20	18
	Distrust. (A1)	.75	.82	11.2	10.5	3.63	3.79	4	4	20	20
	Manip. (A2)	.80	.76	7.62	7.33	3.67	3.36	4	4	20	17
	Women			6.99	6.3	3.30	2.76	4	4	20	17
	Men			8.65	8.61	4	3.61	4	4	19	17
	Opp. (A4)	.71	.66	7.64	7.15	2.84	2.43	4	4	18	17
	Rash. (C6)	.77	.73	8.24	7.89	3.5	3.11	4	4	20	18
BFI-Fr	Total	.74	.76	152	156	15.3	15.2	108	119	197	197
	Extraversion	.87	.88	24.3	24.9	7.32	7.16	9	10	40	40
	Agreeableness	.76	.74	37.9	39.7	6.1	5.53	17	16	50	50
	Conscientiousness	.84	.83	30.8	32.2	6.84	6.22	9	18	45	45
	Neuroticism	.87	.89	25.3	23.7	7.5	7.72	8	8	40	38
	Openness	.77	.81	34.1	36	7.1	6.97	17	15	49	50
MSI-BPD	Total	.71	.65	2.57	2.15	2.24	1.95	0	0	10	8
BSL-23	Total	.92	.94	16.3	11.8	14.1	13.4	0	0	72	57

S1 = Sample 1 and S2 = Sample 2, FFBI-SF = Five-Factor Borderline Inventory, Short Form; Anx. Unc. (N1) = Anxious Uncertainty (Neuroticism 1); Dys. Ang. (N2) = Dysregulated Anger (Neuroticism 2); Desp. (N3) = Despondence (Neuroticism 3); Self-Dist. (N4) = Self-Disturbance (Neuroticism 4); Behav. Dys. (N5) = Behavioral Dysregulation (Neuroticism 5); Aff. Dys. (N6a) = Affective Dysregulation (Neuroticism 6a); Frag. (N6b) = Fragility (Neuroticism 6b); Diss. Tend. (O1) = Dissociative Tendencies (Openness 1); Distrust. (A1) = Distrustfulness (Agreeableness 1); Manip. (A2) = Manipulativeness (Agreeableness 2); Opp. (A4) = Oppositional (Agreeableness 4); Rash. (C6) = Rashness (Conscientiousness 6); MSI-BPD = McLean Screening Instrument for Borderline Personality Disorders; BSL-23 = Borderline Symptom List, 23 items

useless and worthless"); Self-Disturbance (items 4, 16, 28, 40; e.g., "*I can be so different with people that I wonder who I am*"); Behavior Dysregulation (items 5, 17, 29, 41; e.g., "*When I'm upset or edgy, I often do things that later cause me problems*"); Affective Dysregulation (items 6, 18, 30, 42; e.g., "*My emotions can get out of control*"); Fragility (items 7, 19, 31, 43; e.g., "*Even minor failures can cause me great disarray in my life*"); Dissociative Tendencies (items 8, 20, 32, 44; e.g., "*I sometimes feel like I am no longer connected to my body*"); Distrustfulness (items 9, 21, 33, 45; e.g., "*I am unable to trust some of closest friends*"); Manipulativeness (items 10, 22, 34, 46; e.g., "*Sometimes, you have to be manipulative and dishonest to get what you need*"); Oppositional (items 11, 23, 35, 47R; e.g., "*I often conflict with people that are close to me*"); and Rashness (items 12, 24, 36, 48; e.g., "*Others say about me that I do not think before acting*"). One item is reversed. The French short-form version was translated from English to French by two experts in borderline personality disorders, fluent in both French and English reviewed. Then, the translated version was back-translated from French to English. Lastly, the back-translation version was controlled and approved by the original scale's principal author, H. DeShong. After each step, all items were examined by two independent psychiatrists and a bilingual expert.

Big Five Inventory The BFI is a 45-item self-report questionnaire scale assessing five domains of personality (i.e., extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience) rated on a five-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) (John & Srivastava, 1999; French version: Plaisant et al., 2010). The BFI is a shorter version of the NEO Personality Inventory-Revised (NEOPI-R; Costa Jr & McCrae, 1992). Seventeen items are reversed. Coefficient alphas are ranged from .74 to .89 for all domains (see Table 1).

McLean Screening Instrument for Borderline Personality Disorder The MSI-BPD (Zanarini et al., 2003; French version: Mirkovic et al., 2020) is a 10-item self-report scale measuring borderline personality disorder. This True-False questionnaire is based on a subset of questions focusing on the Diagnostic Interview's borderline module for DSM-IV personality disorders (see Zanarini et al., 1996). A total score is calculated by the addition of each item, ranging from 0 to 10. Zanarini et al. (2003) settled a cutoff score at ≥7 (coefficient $\alpha = .71/.65$, see Table 1).

Borderline Symptom List, Short Form The BSL-23 is a 23-item self-rating instrument used to measure specific symptoms of borderline personality disorder (Bohus et al., 2001; Bohus et al., 2009; French version: Nicastro et al., 2016). Each item is answered on a 5-point Likert scale, ranging from 0 (*not at all*) to 4 (*very strong*). A total score is calculated by the addition of each item (coefficient $\alpha \geq .92$, see Table 1)

Statistical Analysis

Given significant differences in FFBI-SF domains between samples 1 and 2 (i.e., for Anxious Uncertainty, Despondence, Self-Disturbance, Affective Dysregulation, Fragility, Dissociative Tendencies, Distrustfulness, and FFBI Total), separate analyses were performed on both samples.

Descriptive analyses were used to describe the demographic characteristics of the samples. Reliability assessments were done using Cronbach's alpha statistics, and the split-half method was applied. M/ANOVA and correlation analyses were performed. Multiple regression analyses were run to determine the incremental validity of the FFBI-SF domains over BSL-23 or MSI-BPD. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were performed using a *maximum likelihood* extraction method combined with a *promax* rotation. The analyses were run on JAMOVI computer software, version 1.6.3 (The Jamovi Project, 2019). Find the raw data file by using the following DOI link: 10.17605/OSF.IO/D3GJH.

Results

Demographic Data

Sample 1 ($N = 335$; women = 208): Age was ranged from 18 to 38 ($M = 21.9$; $SD = 3.12$). Most participants had no children ($n = 330$; 98.5%) and had, on average, one brother and one sister. The main marital status of participants was single (95%) and the main parental socioeconomic status can be described as "*on average*" (51%). Students were predominantly university ($n = 276$; 82%) and Bachelor students ($n = 212$; 63%). Others were master or Ph.D. students ($n = 118$; 35%) or had alternative fields of education ($n = 6$; 2%).

Sample 2 ($N = 162$; women = 90): Age was ranged from 19 to 54 ($M = 31.7$; $SD = 7.52$). Most participants had no children ($n = 121$; 75%) and had, on average, one brother and one sister. The main marital status of participants was single (68.5%) and the main personal socioeconomic status can be described as "*on average*" or "*higher than average*" (76.5%). Participants were predominantly full-time workers ($n = 117$; 72%) and mainly had a Bachelor ($n = 37$; 23%) or Master degree ($n = 92$; 57%).

Table 2 Correlation table

	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1. FFBI-SF	–	–																												
2. BSL-23	.69	.73	–	–																										
3. MSI-BPD	.78	.73	.70	.70	–	–																								
4. Anx. Unc. (N1)	.64	.70	.45	.54	.46	.49	–	–																						
5. Dys. Ang. (N2)	.73	.66	.37	.34	.55	.48	.39	.35	–	–																				
6. Desp. (N3)	.69	.69	.66	.64	.61	.57	.56	.55	.32	.24	–	–																		
7. Self-Dist. (N4)	.72	.71	.58	.49	.61	.48	.37	.42	.37	.29	.56	.54	–	–																
8. Behav. Dys. (N5)	.78	.82	.45	.62	.54	.58	.36	.47	.60	.58	.36	.46	.48	.51	–	–														
9. Aff. Dys. (N6a)	.84	.84	.55	.63	.63	.62	.61	.59	.68	.60	.53	.58	.53	.52	.64	.70	–	–												
10. Frag. (N6b)	.77	.83	.68	.74	.68	.66	.58	.61	.41	.40	.72	.73	.56	.60	.53	.64	.63	.70	–	–										
11. Diss. Tend. (O1)	.61	.67	.57	.62	.63	.53	.29	.45	.25	.28	.54	.54	.46	.48	.39	.48	.38	.51	.56	.63	–	–								
12. Distrust. (A1)	.65	.70	.43	.46	.56	.54	.43	.48	.44	.39	.38	.43	.54	.61	.37	.46	.48	.45	.43	.55	.38	.38	–	–						
13. Manip. (A2)	.54	.52	.25	.32	.31	.36	.07	.17	.36	.24	.16	.13	.37	.37	.49	.47	.33	.30	.25	.32	.27	.30	.32	.39	–	–				
14. Opp. (A4)	.61	.55	.33	.33	.44	.36	.26	.23	.58	.58	.24	.23	.30	.19	.51	.51	.47	.44	.31	.37	.21	.16	.38	.34	.48	.33	–	–		
15. Rash. (C6)	.63	.66	.29	.37	.37	.40	.26	.35	.55	.55	.22	.27	.33	.33	.71	.68	.54	.54	.34	.34	.24	.34	.22	.37	.40	.43	.46	.46	–	–
NEO domain*	–	–	–	–	–	–	.81	.85	.48	.38	.62	.58	.45	.39	.39	.48	.66	.65	.61	.66	.21	.06	−.39	−.32	−.49	−.34	−.69	−.53	−.28	−.24

The left numbers are for Sample 1 and the right numbers for Sample 2. FFBI-SF = Five-Factor Borderline Inventory, Short Form; BSL-23 = Borderline Symptom List, 23 items; MSI-BPD = McLean Screening Instrument for Borderline Personality Disorders; Anx. Unc. (N1) = Anxious Uncertainty (Neuroticism 1); Dys. Ang. (N2) = Dysregulated Anger (Neuroticism 2); Desp. (N3) = Despondence (Neuroticism 3); Self-Dist. (N4) = Self-Disturbance (Neuroticism 4); Behav. Dys. (N5) = Behavioral Dysregulation (Neuroticism 5); Aff. Dys. (N6a) = Affective Dysregulation (Neuroticism 6a); Frag. (N6b) = Fragility (Neuroticism 6b); Diss. Tend. (O1) = Dissociative Tendencies (Openness 1); Distrust. (A1) = Distrustfulness (Agreeableness 1); Manip. (A2) = Manipulativeness (Agreeableness 2); Opp. (A4) = Oppositional (Agreeableness 4); Rash. (C6) = Rashness (Conscientiousness 6). * Corresponding Big Five Inventory domain (John & Srivastava, 1999) for each FFBI-SF subscale

Table 3 Incremental validity table of the FFBI-SF domains

					MSI-BPD					
	R^2		ΔR^2		F		df		p	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Step 1:										
BSL-23	.48	.49	-	-	311.2	154.5	1,333	1,160	< .001	< .001
Step 2:										
FFBI-SF domains	.71	.62	.23	.13	60.7	18.5	13,321	13,148	< .001	< .001
					BSL-23					
Step 1:										
MSI-BPD	.48	.49	-	-	311.2	154.5	1,333	1,160	< .001	< .001
Step 2:										
FFBI-SF domains	.61	.68	.13	.19	38.7	23.9	13,321	13,148	< .001	< .001

Note. This table displays the incremental validity of the FFBI-SF domains over BSL-23 (upper part of the table) in predicting MSI-BPD within both samples and over MSI-BPD (lower part of the table) in predicting BSL-23 within both samples. The left numbers are for Sample 1 and the right numbers for Sample 2. *MSI-BPD* = McLean Screening Instrument for Borderline Personality Disorders; *BSL-23* = Borderline Symptom List, 23-item version; *BFI* = Big Five Inventory; *FFBI-SF* = Five-Factor Borderline Inventory, Short Form

Gender, Age, and Samples Differences in FFBI Domains

For both samples, age was not significantly related to the FFBI domains ($r < .15$) but the same gender differences were found in FFBI domains (S1: $\Lambda = .763$; $F_{(12,322)} = 8.32$; $p < .001$; S2: $\Lambda = .676$; $F_{(12,149)} = 5.96$; $p < .001$). Univariate tests showed significant higher scores for women than men in Anxious Uncertainty (S1: $F_{(1,333)} = 41.5$; $p < .001$; S2: $F_{(1,160)} = 16.7$; $p < .001$) and Affective Dysregulation ($F_{(1,160)} = 4.62$; $p = .03$), and significant higher scores for men in Manipulativeness ($F_{(1,160)} = 21.3$; $p < .001$) in comparison with women (see Table 1).

Psychometric Properties

The internal consistency for both Sample 1 and 2 were excellent, the overall Cronbach's coefficients alpha were .95 (see Table 1) and the split-half method revealed excellent coefficients (> .90). All subscales displayed from satisfactory to good internal consistency and no or little differences were found for both samples. Interdomain correlations were found to be significant and positive. Indeed, on average, all domains were moderatetely (Dancey & Reidy, 2007) intercorrelated ($r_{S1} = .42$; $r_{S2} = .44$). The French version of the FFBI-SF displayed moderate to strong significant correlations with the Five Factor Model (Costa Jr & McCrae, 1992; see Table 2). Anxious Uncertainty ($r_{S1} = .81$; $r_{S2} = .85$), Dysregulated Anger ($r_{S1} = .48$; $r_{S2} = .38$), Despondence ($r_{S1} = .62$; $r_{S2} = .58$), Self-Disturbance ($r_{S1} = .45$; $r_{S2} = .39$), Behavior Dysregulation ($r_{S1} = .39$; $r_{S2} = .48$), Affective Dysregulation ($r_{S1} = .66$; $r_{S2} = .65$) and Fragility ($r_{S1} = .61$; $r_{S2} = .66$) were significantly correlated with neuroticism; Distrustfulness ($r_{S1} = -.39$; $r_{S2} = -.32$), Manipulativeness ($r_{S1} = -.49$; $r_{S2} = -.34$) and Oppositional ($r_{S1} = -.69$; $r_{S2} = -.53$) showed significant moderate to strong correlations with agreeableness; Rashness was significantly related with conscientiousness ($r_{S1} = -.28$; $r_{S2} = -.24$); lastly, Dissociative Tendencies was significantly correlated with openness to experience ($r_{S1} = .21$; no significant correlation was found for S2: $r = .06$).

Convergent Validity

FFBI Total scores had strong and significant correlations with BSL-23 ($r_{S1} = .69$; $r_{S2} = .73$) and MSI-BPD ($r_{S1} = .78$; $r_{S2} = .73$). Table 2 shows the convergent validity of the 12 FFBI-SF subscales with the BSL-23 and the MSI-BPD.

Incremental Validity

The incremental validity table (Table 3) provides parts of the variance explained by each variable (or block of variables) in predicting MSI-BPD or BSL-23 scores. Results show that the FFBI-SF domains accounted for significant additional parts of variance (13 to 23%) in predicting MSI-BPD scores over BSL-23, and in predicting the BSL-23 scores over MSI-BPD (13 to 19%).

Table 4 CFA factor loadings of the french version of FFBI-SF

Items	AU		DA		D		SD		BD		AD		F		DT		Di		M		O		R	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1. I tend to be quite anxious. | *J'ai tendance à être assez anxieux(se).*	.95	.96																						
2. I have already had some angry outbursts. | *J'ai déjà fait quelques crises de colère.*			.99	.92																				
3. I sometimes feel useless and worthless. | *Je me sens parfois inutile et sans valeur.*					.87	1.1																		
4. I can be so different according to people, like I am not the same person. | *Je peux être tellement différent en fonction des personnes, comme si je n'étais pas la même personne.*							.66	.68																
5. I frequently have urges to do things that cause me trouble. | *J'ai fréquemment des envies irrépressibles de faire des choses qui me causent des ennuis.*									.65	.54														
6. My emotions can get out of control. | *Mes émotions peuvent devenir incontrôlables.*											1.1	.89												
7. Hurting myself is one of the few ways that I can tolerate my emotions. | *Me faire du mal est l'une des seules façons qui me permette de tolérer mes émotions.*													.53	.28										
8. I have already felt that things were unreal and that I was detached from life (ex: feeling disconnected from my body). | *J'ai déjà ressenti que des choses étaient irréelles et que j'étais détaché(e) de la vie (ex: impression d'être déconnecté(e) de mon corps).*															1.1	1.1								
9. I often distrust others. | *Je me méfie souvent des autres.*																	.67	.78						
10. I sometimes do things I shouldn't do to obtain from others things I want or need. | *Je fais parfois des choses que je ne devrais pas pour obtenir des autres qu'ils fassent ce dont j'ai envie ou besoin.*																			.89	.73				
11. I frequently get into conflict with others. | *J'ai tendance à entrer fréquemment en conflit.*																					.83	.51		
12. I get into trouble because I don't anticipate the consequences of my own actions. | *Je m'attire des problèmes car je n'anticipe pas les conséquences de mes actions.*																							.72	.56
13. I worry a lot. | *Je m'inquiète énormément.*	1.1	1.1																						
14. My anger often seems to me out of control. | *Ma colère me semble souvent incontrôlable.*			1.0	.84																				
15. I have thought about ways to commit suicide. | *Il m'est arrivé de réfléchir à des moyens de me suicider.*					1.0	.94																		
16. I can be so different with people that I wonder who I am. | *Je peux être si différent(e) en fonction des personnes que je me demande qui je suis.*							.89	.59																
17. Sometimes, I get carried away by my urges. | *Parfois, je me laisse emporter par mes envies irrépressibles.*									.79	.76														
18. I feel like I don't have a lot of control over how I feel. | *J'ai le sentiment de ne pas avoir beaucoup de contrôle sur ce que je ressens.*											.90	.76												
19. I have already threatened to kill myself. | *J'ai déjà menacé de me suicider.*													.55	.45										
20. I sometimes feel like I am no longer connected to my body. | *J'ai parfois l'impression de ne plus être connecté(e) à mon corps.*															1.0	.96								
																	.73	.86						

Table 4 (continued)

Items	AU		DA		D		SD		BD		AD		F		DT		Di		M		O		R	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
21. It is really difficult for me to trust people. | *C'est vraiment difficile pour moi de faire confiance aux gens.*																								
22. People have already qualified me as a manipulator. | *Des personnes m'ont déjà qualifié de manipulateur(trice).*																			.82	.56				
23. I would threaten people to make them do what I want. | *Je menacerais des gens pour les obliger à faire ce que je veux.*																					.24	.27		
24. I tend to act quickly without asking myself questions. | *J'ai tendance à agir rapidement sans me poser de questions.*																							.81	.63
25. I am very afraid about people leaving me or forsaking me. | *J'ai très peur que les gens me quittent ou me délaissent.*	.70	.74																						
26. My anger sometimes gets the better of me. | *Ma colère prend parfois le dessus sur moi.*			1.2	1.1																				
27. I often feel sad. | *Je me sens souvent triste.*					.88	.98																		
28. I feel like I am not close to anyone. | *J'ai l'impression de n'être proche de personne.*							.98	.94																
29. When I'm upset or edgy, I often do things that later cause me problems. | *Quand je suis contrarié(e) ou énervé(e), je fais souvent des choses qui me posent plus tard des problèmes.*									.82	.88														
30. My mood quickly changes from one feeling to another. | *Mon humeur bascule rapidement d'un sentiment à l'autre.*											.97	1.0												
31. Even minor failures can cause me great disarray in my life. | *Des échecs, même mineurs, peuvent me causer un grand désarroi dans ma vie.*													.77	1.0										
32. Sometimes, I feel like I am not real. | *J'ai parfois l'impression de ne pas être réel.*															.99	.69								
33. People are not as faithful and loyal to me as I would like. | *Les gens ne me sont pas aussi fidèles et loyaux que je le voudrais.*																	.76	.94						
34. I am known for hiding or embellishing the truth in order to reach my goals. | *Je suis connu pour dissimuler ou embellir la vérité afin d'arriver à mes fins.*																			.71	.82				
35. I often conflict with people that are close to me. | *Je rentre souvent en conflit avec les personnes qui me sont proches.*																					.86	.84		
36. Others say about me that I do not think before acting. | *Les autres disent de moi que je ne réfléchis pas avant d'agir.*																							.81	.73
37. I worry a lot about things that are out of my control. | *Je m'inquiète énormément pour les choses qui sont hors de mon contrôle.*	.97	1.0																						
38. My anger sometimes caused me trouble. | *Ma colère m'a parfois causé des ennuis.*			1.0	.98																				
39. I have been thinking about suicide since I was a teenager. | *Je pense au suicide depuis mon adolescence.*					.88	.73																		
40. I often feel like excluded or rejected by others. | *Je me sens souvent comme exclu(e) ou rejeté(e) par les autres.*							.93	.95																
41. I have already done many things impulsively that I later regret. | *J'ai déjà fait beaucoup de choses impulsivement que j'ai regrettées ensuite.*									.93	.97														
											1.1	1.0												

Table 4 (continued)

Items	AU		DA		D		SD		BD		AD		F		DT		Di		M		O		R	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
42. I have a hard time controlling my moods. | *J'ai du mal à contrôler mes humeurs.*																								
43. I don't think I can go on living this way. | *Je ne pense pas pouvoir continuer à vivre de cette façon.*													.80	.68										
44. Sometimes, I feel like nothing is real. | *J'ai parfois l'impression que rien n'est réel.*															.93	.59								
45. I am unable to trust some of closest friends. | *Je ne parviens pas à faire confiance à certain(e)s de mes ami(e)s les plus proches.*																	.95	.83						
46. Sometimes, you have to be manipulative and dishonest to get what you need. | *Parfois, il faut être manipulateur(trice) et malhonnête pour obtenir ce dont on a besoin.*																			.85	.83				
47. I am easy to get along with. | *Je suis facile à vivre.*																					.55	.38		
48. I have done very bad things under the influence of impulsivity. | *J'ai déjà fait de très mauvaises choses sous le coup de l'impulsivité.*																							.76	.72

The left numbers are for Sample 1 and the right numbers for Sample 2. AU = Anxious Uncertainty; DA = Dysregulated Anger; D = Despondence; SD = Self-Disturbance; BD = Behavioral Dysregulation; AD = Affective Dysregulation; F = Fragility; DT = Dissociative Tendencies; Di = Distrustfulness; M = Manipulativeness; O = Oppositional; R = Rashness

Factor Analysis

The overall Kaiser-Meyer-Olkin (KMO) measure was very high (S1 = .91; S2 = .87) and the Bartlett's test of sphericity was significant for both samples (S1: $\chi^2_{(1128)} = 9342$; $p < .001$; S2: $\chi^2_{(1128)} = 5102$; $p < .001$). The model fit measures from the EFA showed a significant model (S1: $\chi^2_{(618)} = 953$; $p < .001$; S2: $\chi^2_{(618)} = 799$; $p < .001$), the Root Mean Square Error of Approximation (RMSEA) value (S1 = .0447; S2 = .0573) revealed a good fit model, and the Tucker Lewis Index (TLI) indicated that the model improves the fit by over 91% relative to the null model (S1: *TLI* = .923; S2: *TLI* = .911). The 12-factor model explained ≥60% of the variance (S1: 60%; S2: 64%).

The CFA test for exact fit was significant for both samples (S1: $\chi^2_{(1014)} = 2561$; $p < .001$; $\chi^2/Df = 2.53$; S2: $\chi^2_{(1014)} = 2095$; $p < .001$; $\chi^2/Df = 2.07$). The Standardized Root Mean Square Residual (SRMR; S1 = .0636; S2 = .0785), RMSEA (S1 = .0675; S2 = .0811), Comparative Fit Index (CFI; S1 = .823; S2 = .765) and TLI (S1 = .803; S2 = .739) revealed that the 12-factor model is an acceptable fit model (see Table 4 for factor loadings).[3] Item 23 ("*I would threaten people to make them do what I want")* presented low loadings on the Oppositional domain. To solve this problem, the EFA analysis revealed that item 23 would have a better fit with the Manipulativeness domain. Otherwise, removing the item would be another valid solution.

Note that a cutoff score of the FFBI-SF can be determined for clinical purposes. Taking both samples together ($N = 497$; $M = 111$; $SD = 31.1$), with a set-out threshold of .05 in one tail (1.645) and giving a simple modification of the Z score formula ($X = M \pm 1.96 \times SD$), we found the following cut-off score: $111 + 1.645 \times 31.1 = \mathbf{162}$. Therefore, people scoring ≥162 in total FFBI-SF can be considered as presenting salient BPD traits (concomitantly, clinical investigations are still required to substantiate and confirm a diagnosis).

Discussion

This study aimed to investigate the psychometric properties of the French version of the FFBI-SF in the general population. Our results showed that this French adaptation reaches good psychometric properties in both samples collected. Excellent coefficients alpha were found for the whole scale and with the split-half method. Furthermore, each FFBI domain displayed from acceptable to good internal consistency, except for the Fragility domain that presented lower internal consistency. All domains together, the average Cronbach's alpha is .78, showing reliable domains. The FFBI-SF across-domain correlations displayed that all domains were significantly and positively interrelated, suggesting that these domains are concordant.

Concerning the personality measurement (i.e., the BFI, a tool based on the FFM), our findings revealed that the correlations between FFBI-SF domains and their corresponding FFM facets were close to those reported by DeShong et al. (2016). These results are in line with Mullins-Sweatt et al.'s (2012, p.483) statement: "*the FFBI subscales were constructed to provide an assessment of borderline maladaptive variants of each respective NEO PI-R facet."* Furthermore, the convergent analysis showed strong correlations between the French version of the FFBI-SF and the BSL-23 ($r \geq .69$) and the MSI-BPD ($r \geq .73$), showing that those scales assess a similar construct. Similar results were found by Helle et al. (2018) between the FFBI-SF and the MSI-BPD. The incremental validity analyses revealed that the FFBI-SF domains provided a substantial additional part of variance (13 to 23%) in predicting the MSI-BPD or BSL-23. Similarly, Mullins-Sweatt et al. (2012) found a comparable incremental validity when predicting the borderline subscale of the Personality Diagnostic Questionnaire-4 (Bagby & Farvolden, 2004), the Schedule for Nonadaptive and Adaptive Personality (Clark, 1993), or the Wisconsin Personality Disorder Inventory (Klein et al., 1993).

Lastly, our findings from the EFA and CFA suggest that the 12-factor model proposed by Mullins-Sweatt et al. (2012) and DeShong et al. (2016) gives, overall, an acceptable fit to our data. Indeed, the relative chi-square (χ^2/Df) is required to be less than 2 to less than 5 (Schumacker & Lomax, 2004; Ullman, 2001) to display an acceptable model and our data fit with this requirement. Furthermore, following Steiger's (2007) or Hu and Bentler's (1999) recommendations, RMSEA and SRMR indices below or equal .07 (for the former) and .08 (for the latter) reveal an acceptable model, as found in our results. Lastly, the Comparative Fit Index and the Tucker Lewis Index (CFI and TLI) from the EFA were ≥ .91, showing an acceptable model fit. In contrast, the CFI and TLI indices from the CFA showed lower values than required (≥.95; Hu & Bentler, 1999). Item 23 was found to display low loadings on the Oppositional domain, and, as a solution, we suggest associating this item with the Manipulativeness domain or removing it.

In conclusion, this study is inserted in a growing need for researchers and therapists to benefit from reliable and valid existing tools that assess borderline personality disorders in different languages. This study's overall results suggest that the French version of the FFBI-SF has acceptable model fit indices. It is a reliable and valid instrument, quickly fulfilled (around five minutes), for measuring borderline traits among clinical and non-clinical populations. Moreover, the FFBI-SF scale provides the advantage to display a multifactorial

[3] We also run the CFA on the whole sample (S1 + S2) and found quite similar fit indices (χ^2/Df = 3.14; *SRMR* = .0591; *RMSEA* = .0656; *CFI* = .832; *TLI* = .813).

approach, allowing researchers or therapists to precisely detect pathological traits among 12 domains. As displayed, the incremental validity showed that the FFBI-SF brought a significant contribution to BPD assessment. Lastly, for clinical purposes and concomitantly to clinical investigations, a cutoff score has been determined (≥ 162) to identify people with salient BPD traits (i.e., subclinical BPD).

Limitations and Future Directions

Some limitations must be considered in this study. As a self-report questionnaire, several biases may be present (e.g., anosognosia or social desirability). Furthermore, the FFBI-SF scale is a relatively recent measure. Therefore, cumulative works on independent, diverse and robust samples are still needed to sustain its psychometric properties empirically.

As future directions for studies, the French version of FFBI-SF should also be tested among a BPD clinical population, notably to test its discriminant validity. As DeShong et al. (2016) reported, the FFBI presented strong discriminant validity; comparable results should be expected. In addition, it would be interesting to test different clinical populations (in- or outpatients) to determine the prevalence of borderline traits among various psychopathologies. Lastly, we encourage FFBI-SF translation and adaptation in several languages to establish, for instance, potential influences of culture on borderline personality traits.

Acknowledgments The authors would like to thank Dr. Hilary DeShong for her precious help.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The ethical committee of the Department of Psychology of the University of Liège (Belgium) approved the study, reference n°: 1920–92.

Informed Consent Informed consent was obtained from all participants included in the study.

Online Data The online data file is available on the following DOI link: https://doi.org/10.17605/OSF.IO/D3GJH.

References

Alfonso, G., Franco, B. C., Cervigni, M., Martino, P., Paolasini, M. G. V., Gargiulo, A. P. I., Gargiulo, A. J. M., & Gargiulo, P. A. (2021). In P. A. Gargiulo, & H. L. M. Arroyo (Eds.). *Psychiatry and neuroscience update. From epistemology to clinical psychiatry – Vol. IV* (pp. 527-536). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-61721-9_37.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

Bagby, R. M., & Farvolden, P. (2004). Comprehensive handbook of psychological assessment: Vol. personality assessment. In Hilsenroth, M. J., Segal, D. L., & Hersen, M. (Eds.), *The personality diagnostic questionnaire–4 (PDQ–4)*. (Vol. 2, 122–133). Wiley.

Bender, D. S., Dolan, R. T., Skodol, A. E., Sanislow, C. A., Dyck, I. R., McGlashan, T. H., Shea, M. T., Zanarini, M. C., Oldham, J. M., & Gunderson, J. G. (2001). Treatment utilization by patients with personality disorders. *American Journal of Psychiatry, 158*(2), 295–302. https://doi.org/10.1176/appi.ajp.158.2.295.

Bohus, M., Limberger, M. F., Frank, U., Sender, I., Gratwohl, T., & Stieglitz, R. D. (2001). Entwicklung der borderline-symptom-liste. *Psychotherapie Psychosomatik Medizinische Psychologie, 51*(5), 201–211. https://doi.org/10.1055/s-2001-13281.

Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R. D., Domsalla, M., Chapman, A. L., Steil, R., Philipsen, A., & Wolf, M. (2009). The short version of the borderline symptom list (BSL-23): Development and initial data on psychometric properties. *Psychopathology, 42*(1), 32–39. https://doi.org/10.1159/000173701.

Carvalho, F. L., Pianowski, G., Bacciotti, J., & Reis, A. M. (2018). Assessing borderline personality disorder based on the hierarchical taxonomy of psychopathology (HiTOP): Dimensional clinical personality inventory 2 – BPD. *Archives of Psychiatry & Psychotherapy, 20*(4), 77–87. https://doi.org/10.12740/APP/99146.

Clark, L. A. (1993). *Manual for the schedule for nonadaptive and adaptive personality*. University of Minnesota Press.

Costa Jr., P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*(6), 653–665. https://doi.org/10.1016/0191-8869(92)90236-I.

Costa Jr., P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64*(1), 21–50. https://doi.org/10.1207/s15327752jpa6401_2.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Pearson education.

DeShong, H. L., Mullins-Sweatt, S. N., Miller, J. D., Widiger, T. A., & Lynam, D. R. (2016). Development of a short form of the five-factor borderline inventory. *Assessment, 23*(3), 342–352. https://doi.org/10.1177/1073191115581475.

Gunderson, J. G. (2011). Clinical practice: Borderline personality disorder. *The New England Journal of Medicine, 364*(21), 2037–2042. https://doi.org/10.1056/NEJMcp1007358.

Ha, C., Balderas, J. C., Zanarini, M. C., Oldham, J., & Sharp, C. (2014). Psychiatric comorbidity in hospitalized adolescents with borderline personality disorder. *Journal of Clinical Psychiatry, 75*, e457–e464.

Helle, A. C., DeShong, H. L., Baraldi, A. N., & Mullins-Sweatt, S. N. (2018). The five factor borderline inventory: Behavioral outcomes across time. *Psychological Assessment, 30*(1), 43–61. https://doi.org/10.1037/pas0000502.

Hopwood, C. J., Kotov, R., Krueger, R. F., Watson, D., Widiger, T. A., Althoff, R. R., Ansell, E. B., Bach, B., Michael Bagby, R., Blais, M. A., Bornovalova, M. A., Chmielewski, M., Cicero, D. C., Conway, C., de Clercq, B., de Fruyt, F., Docherty, A. R., Eaton, N. R., Edens, J. F., Forbes, M. K., Forbush, K. T., Hengartner, M. P., Ivanova, M. Y., Leising, D., John Livesley, W., Lukowitsky, M. R., Lynam, D.

R., Markon, K. E., Miller, J. D., Morey, L. C., Mullins-Sweatt, S. N., Hans Ormel, J., Patrick, C. J., Pincus, A. L., Ruggero, C., Samuel, D. B., Sellbom, M., Slade, T., Tackett, J. L., Thomas, K. M., Trull, T. J., Vachon, D. D., Waldman, I. D., Waszczuk, M. A., Waugh, M. H., Wright, A. G. C., Yalch, M. M., Zald, D. H., & Zimmermann, J. (2018). The time has come for dimensional personality disorder diagnosis. *Personality and Mental Health, 12*(1), 82–86. https://doi.org/10.1002/pmh.1408.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

John, O. P., & Srivastava, S. (1999). The big-five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin& O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol.2, pp. 102–138). Guilford press.

Klein, M. H., Benjamin, L. S., Rosenfeld, R., Treece, C., Husted, J., & Greist, J. H. (1993). The Wisconsin personality disorders inventory: Development, reliability, and validity. *Journal of Personality Disorders, 7*(4), 285–303. https://doi.org/10.1521/pedi.1993.7.4.285.

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., et al. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126*(4), 454–477. https://doi.org/10.1037/abn0000258.

Leichsenring, F., Leibing, E., Kruse, J., New, A. S., & Leweke, F. (2011). Borderline personality disorder. *Lancet, 377*(9759), 74–84. https://doi.org/10.1016/S0140-6736(10)61422-5.

Miller, J. D., Morse, J. Q., Nolf, K., Stepp, S. D., & Pilkonis, P. A. (2012). Can DSM-IV borderline personality disorder be diagnosed via dimensional personality traits? Implications for the DSM-5 personality disorder proposal. *Journal of Abnormal Psychology, 121*(4), 944–950. https://doi.org/10.1037/a0027410.

Mirkovic, B., Speranza, M., Caihol, L., Guelfi, J.-D., Perez-Diaz, F., Corcos, M., Robin, M., & Pham-Scottez, A. (2020). Validation of the French version of the McLean screening instrument for borderline personality disorder (MSI-BPD) in an adolescent sample. *BMC Psychiatry, 20*, 222. https://doi.org/10.1186/s12888-020-02643-8.

Mullins-Sweatt, S. N., Edmundson, M., Sauer-Zavala, S., Lynam, D. R., Miller, J. D., & Widiger, T. A. (2012). Five-factor measure of borderline personality traits. *Journal of Personality Assessment, 94*(5), 475–487. https://doi.org/10.1080/00223891.2012.672504.

Nicastro, R., Prada, P., Kung, A. L., Salamin, V., Dayer, A., Aubry, J. M., Guenot, F., & Perroud, N. (2016). Psychometric properties of the French borderline symptom list, short form (BSL-23). *Borderline Personality Disorder and Emotion Dysregulation, 3*(1), 4. https://doi.org/10.1186/s40479-016-0038-0.

Plaisant, O., Courtois, R., Réveillère, C., Mendelsohn, G. A., & John, O. P. (2010). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. *Annales Médico-Psychologiques, Revue Psychiatrique, 168*(2), 97–106. https://doi.org/10.1016/j.amp.2009.09.003.

Samuel, D. B., & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review, 28*(8), 1326–1342. https://doi.org/10.1016/j.cpr.2008.07.002.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (Second ed.). Lawrence Erlbaum Associates.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017.

The Jamovi Project (2019). Jamovi, version 1.0 [Computer Software]. Retrieved from https://www.jamovi.org/

Trull, T. J., & Brown, W. C. (2013). Borderline personality disorder: A five-factor model perspective. In T. A. Widiger & P. T. Costa, Jr. (Eds.), *Personality disorders and the five-factor model of personality* (p. 119–132). American Psychological Association. https://doi.org/10.1037/13939-008.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (2001). *Using Multivariate Statistics* (4th ed& pp 653–771). : Allyn & Bacon.

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist, 62*(2), 71–83. https://doi.org/10.1037/0003-066X.62.2.71.

Zanarini, M. C., Frankenburg, F. R., Sickel, A. E., & Yong, L. (1996). *The diagnostic interview for DSM-IV personality disorders (DIPD-IV)*. McLean Hospital.

Zanarini, M. C., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. A. (2003). Screening measure for BPD: The McLean screening instrument for borderline personality disorder (MSI-BPD). *Journal of Personality Disorders, 17*(6), 568–573. https://doi.org/10.1521/pedi.17.6.568.25355.

Zanarini, M. C., Frankenburg, F. R., Reich, D. B., & Fitzmaurice, G. (2012). Attainment and stability of sustained symptomatic remission and recovery among patients with borderline personality disorder and axis II comparison subjects: A 16-year prospective follow-up study. *American Journal of Psychiatry, 169*(5), 476–483. https://doi.org/10.1176/appi.ajp.2011.11101550.

Annexe E : Trolley-like problems

The company's dilemma (French version)		
The boss (P_A)	**The spokesperson (P_B)**	**One of the five employees (P_C)**
Depuis plus de 20 ans, vous êtes la patronne d'une petite entreprise. Vous appréciez votre travail et bénéficiez d'un bon poste ainsi qu'une bonne rémunération. Toutefois, la crise sanitaire liée à la pandémie a mis l'entreprise franchement dans le rouge et une restructuration du personnel semble imminente. Paniqués, les six employés de votre service se sont mobilisés pour faire grève car ils appréhendent des licenciements sans que des indemnités de licenciement soient versées. Ils font donc grève pour revendiquer une sécurité d'emploi et pour obtenir des indemnités de licenciement correctes en cas de restructuration. Pour tenter de résoudre la situation, vous convoquez la porte-parole du mouvement. Vous comprenez la position de cette dernière et vous lui expliquez ne pas avoir énormément d'options pour arranger la situation. Pour vous, il y a deux options à effets immédiats et sans concertation possible. La première option est de garantir contractuellement à la porte-parole, et à elle seule, son emploi. Vous licencierez alors ses cinq autres collègues sans indemnités de licenciement. La seconde option est de licencier la porte-parole, et elle seule, sans indemnités de licenciement. Ses cinq autres collègues auront alors une garantie contractuelle de leur emploi. Dans les deux cas, personne ne saura que cette décision émane de vous. N.B. : il n'y a pas de concertation avec la porte-parole, vous êtes la seule à décider. Si vous choisissez la première option, la porte-parole bénéficiera d'une garantie d'emploi contractuelle. Toutefois, ses cinq autres collègues seront licenciés sans indemnités. Si vous choisissez la seconde option, la porte-parole sera licenciée sans indemnités. Toutefois, ses cinq autres collègues bénéficieront d'une garantie contractuelle de leurs emplois.	Depuis plus de 20 ans, vous êtes employée dans une petite entreprise. Vous appréciez votre travail et bénéficiez d'un bon poste ainsi qu'une bonne rémunération. Toutefois, la crise sanitaire liée à la pandémie a mis l'entreprise franchement dans le rouge et une restructuration du personnel semble imminente. Paniqués, les six employés de votre service (dont vous) se sont mobilisés pour faire grève car vous appréhendez des licenciements sans que des indemnités de licenciement soient versées. Vous faites donc grève pour revendiquer une sécurité d'emploi et pour obtenir des indemnités de licenciement correctes en cas de restructuration. Pour tenter de résoudre la situation, votre patron vous convoque car vous êtes la porte-parole du mouvement. Il comprend votre position et vous explique ne pas avoir énormément d'options pour arranger la situation. Votre patron vous propose un arrangement (seuls vous et lui serez au courant de ce dernier). Cet arrangement comporte deux options à effets immédiats et sans concertation possible. La première option est de vous garantir contractuellement à vous, et vous seule, votre emploi. Vos cinq autres collègues seront alors licenciés sans indemnités de licenciement. La seconde option est de vous licencier vous, et vous seule, sans indemnités de licenciement. Vos cinq autres collègues auront alors une garantie contractuelle de leur emploi. Comme mentionné, dans les deux cas, personne ne saura que cette décision émane de vous. N.B. : il n'y a pas de concertation avec le patron, vous êtes la seule à décider. Si vous choisissez la première option, vous bénéficierez d'une garantie d'emploi. Toutefois, votre employeur licenciera vos cinq autres collègues sans indemnités. Si vous choisissez la seconde option, vous serez licenciée sans indemnités. Toutefois, votre employeur garantira l'emploi de vos cinq autres collègues.	Depuis plus de 20 ans, vous êtes employée dans une petite entreprise. Vous appréciez votre travail et bénéficiez d'un bon poste et d'une bonne rémunération. Toutefois, la crise sanitaire liée à la pandémie a mis l'entreprise franchement dans le rouge et une restructuration du personnel semble imminente. Paniqués, les six employés de votre service (dont vous) se sont mobilisés pour faire grève car vous appréhendez des licenciements sans que des indemnités de licenciement soient versées. Vous faites donc grève pour revendiquer une sécurité d'emploi et pour obtenir des indemnités de licenciement correctes en cas de restructuration. Pour tenter de résoudre la situation, votre patron convoque votre porte-parole du mouvement. Il comprend sa position et lui explique ne pas avoir énormément d'options pour arranger la situation. Plus tard, vous croisez votre patron et il se met à discuter avec vous. Il vous propose un arrangement (seuls vous et lui serez au courant de ce dernier). Cet arrangement comporte deux options à effets immédiats et sans concertation possible. La première option est de garantir contractuellement à votre porte-parole, et à elle seule, son emploi. Vos quatre autres collègues et vous serez alors licenciés sans indemnités de licenciement. La seconde option est de licencier votre porte-parole, et elle seule, sans indemnités de licenciement. Vos quatre autres collègues et vous aurez alors une garantie contractuelle de vos emplois. Comme mentionné, dans les deux cas, personne ne saura que cette décision émane de vous. N.B. : il n'y a pas de concertation avec votre patron, vous êtes la seule à décider. Si vous choisissez la première option, votre porte-parole bénéficiera d'une garantie d'emploi contractuelle. Toutefois, vos quatre autres collègues et vous serez licenciés sans indemnités. Si vous choisissez la seconde option, votre porte-parole sera licenciée sans indemnités. Toutefois, vos quatre autres collègues et vous bénéficierez d'une garantie contractuelle de vos emplois.

The early departure dilemma (French version)		
The friend (P_A)	**The homeless (P_B)**	**One of the five roommates (P_C)**
Votre amie doit quitter impérativement son appartement actuel deux semaines avant d'emménager dans le nouveau et elle a besoin d'un endroit où aller. Vous habitez avec quatre autres amis qui sont très stressés pour le moment car ils sont en période d'examens de fin d'année. Ils ne s'entendent pas avec votre amie et ils se sont déjà disputés avec elle par le passé. D'ailleurs, la dernière fois qu'elle est venue, ils en ont eu vraiment marre d'elle. N'ayant pu décrocher, vous recevez un message vocal de votre amie sur votre répondeur. En détresse, elle vous explique sa situation et vous dit qu'elle ne sait pas où aller. Elle prévoit donc de venir s'installer chez vous dès ce soir. Plus tard dans la soirée, votre amie se présente à votre porte et vous lui ouvrez. Sans concertation avec les autres, vous devez choisir parmi deux options. La première option est de laisser votre amie séjourner parmi vous. En choisissant cette option, elle aura un endroit où dormir pendant deux semaines. Cependant, vos quatre autres colocataires devront accueillir votre amie pendant leur période d'examens. La seconde option est de ne pas laisser votre amie séjourner parmi vous. En choisissant cette option, elle n'aura nulle part où dormir pendant deux semaines. Cependant, vos quatre autres colocataires ne devront pas accueillir votre amie pendant leur période d'examens.	Vous devez quitter impérativement votre appartement actuel deux semaines avant d'emménager dans le nouveau et vous avez besoin d'un endroit où aller. Votre amie habite avec quatre autres amis à elle qui sont très stressés pour le moment car ils sont en période d'examens de fin d'année. Vous ne vous entendez pas avec eux et vous vous êtes déjà disputés par le passé. D'ailleurs, la dernière fois que vous êtes venue, ils en ont eu vraiment marre de vous. Vous appelez votre amie et, n'ayant pu décrocher, vous tombez sur son répondeur et lui laissez un message vocal. En détresse, vous lui expliquez votre situation et lui dites que vous ne savez pas où aller. Vous prévoyez donc de venir vous installer chez elle dès ce soir. Plus tard dans la soirée, vous vous présentez à sa porte et votre amie vous ouvre. Sans concertation avec les autres, vous devez choisir parmi deux options. La première option est de vous installer chez votre amie pour séjourner parmi eux. En choisissant cette option, vous aurez un endroit où dormir pendant deux semaines. Cependant, les quatre colocataires de votre amie devront vous accueillir pendant leur période d'examens. La seconde option est de vous raviser et de ne pas séjourner parmi eux. En choisissant cette option, vous n'aurez nulle part où dormir pendant deux semaines. Cependant, les quatre colocataires de votre amie ne devront pas vous accueillir pendant leur période d'examens.	Une amie d'un de vos colocataires doit quitter impérativement son appartement actuel deux semaines avant d'emménager dans le nouveau et elle a besoin d'un endroit où aller. Vous habitez à cinq dans votre habitation et quatre d'entre vous (dont vous) êtes très stressés pour le moment car vous êtes en période d'examens de fin d'année. Vous ne vous entendez pas avec l'amie de votre colocataire et vous vous êtes déjà disputés avec elle par le passé. D'ailleurs, la dernière fois qu'elle est venue, vous en avez eu vraiment marre d'elle. N'ayant pu décrocher, vous recevez un message vocal de votre colocataire sur votre répondeur. Il vous dit que son amie l'a appelée en détresse car elle ne sait pas où aller. Elle prévoit donc de venir s'installer chez vous dès ce soir. Votre colocataire vous demande de choisir à sa place. Plus tard dans la soirée, l'amie de votre colocataire se présente à votre porte et vous lui ouvrez. Sans concertation avec les autres, vous devez choisir parmi deux options. La première option est de laisser l'amie de votre colocataire séjourner parmi vous. En choisissant cette option, elle aura un endroit où dormir pendant deux semaines. Cependant, vos trois autres colocataires et vous devrez l'accueillir pendant votre période d'examens. La seconde option est de ne pas laisser l'amie de votre colocataire séjourner parmi vous. En choisissant cette option, elle n'aura nulle part où dormir pendant deux semaines. Cependant, vos trois autres colocataires et vous ne devrez pas l'accueillir pendant votre période d'examens.

The transplant dilemma (French version)		
The surgeon (P_A)	**The new patient (P_B)**	**One of the five dying patients (P_C)**
Vous êtes une brillante chirurgienne spécialiste des transplantations. Face à la pandémie mondiale, vous êtes confrontée à des choix que jamais auparavant vous n'avez dû faire. Au sein de votre service se trouvent cinq patients nécessitants chacun un organe différent sans lequel il mourra. Malheureusement, aucun organe n'est disponible pour ces transplantations. Arrive alors une autre patiente en bonne santé pour une petite opération bénigne mais urgente. Pendant l'examen, vous constatez que tous ses organes sont compatibles avec les cinq autres patients mourants. Ne sachant quoi faire, vous avez la possibilité de prendre une décision totalement atypique dans votre pratique. Il y a deux options : (1) vous réalisez une manœuvre létale pendant l'opération de la nouvelle patiente. Cette manœuvre causera donc sa mort mais vous n'en serez pas inquiétée (personne ne pourra déterminer que cette action médicale était intentionnelle). Ceci vous permettra de prélever ses organes afin de faire vivre vos cinq autres patients ; (2) vous réalisez l'opération bénigne de la nouvelle patiente dans les règles de l'art. Elle rentrera sous peu chez elle sans que ses organes n'aient été prélevés mais vos cinq autres patients mourront.	Vous êtes en bonne santé et vous vous rendez dans un hôpital pour une petite opération bénigne mais urgente. La docteure qui s'occupe de vous est une brillante chirurgienne spécialiste des transplantations. Face à la pandémie mondiale, elle est confrontée à des choix que jamais auparavant elle n'avait dû faire. Elle vous explique qu'au sein de son service se trouvent cinq patients nécessitants chacun un organe différent sans lequel il mourra. Malheureusement, aucun organe n'est disponible pour ces transplantations. La docteure constate que tous vos organes sont compatibles avec ses cinq autres patients mourants. Ne sachant quoi faire, la docteure se tourne vers vous pour prendre une décision totalement atypique dans sa pratique. Il y a deux options : (1) la docteure réalise une manœuvre létale pendant votre opération. Cette manœuvre causera donc votre mort mais elle n'en sera pas inquiétée (personne ne pourra déterminer que cette action médicale était intentionnelle). Ceci lui permettra de prélever vos organes afin de faire vivre les cinq autres patients ; (2) la docteure réalise votre opération bénigne dans les règles de l'art. Vous rentrerez sous peu chez vous sans que vos organes n'aient été prélevés mais les cinq autres patients mourront.	Vous et quatre autres personnes êtes mourants. En effet, chacun d'entre vous nécessite une transplantation d'un organe différent sans lequel vous mourrez. Malheureusement, aucun organe n'est disponible pour ces transplantations. Une brillante chirurgienne spécialiste des transplantations s'occupe de vous cinq. Face à la pandémie mondiale, elle est confrontée à des choix que jamais auparavant elle n'avait dû faire. Arrive alors une autre patiente en bonne santé pour une petite opération bénigne mais urgente. Pendant l'examen, votre docteure constate que tous les organes de cette nouvelle patiente sont compatibles avec les quatre autres patients mourants et vous. Ne sachant quoi faire, la docteure se tourne vers vous pour rendre une décision totalement atypique dans sa pratique. Il y a deux options : (1) la docteure réalise une manœuvre létale pendant l'opération de cette nouvelle patiente. Cette manœuvre causera donc sa mort mais elle n'en sera pas inquiétée et vous non plus (personne ne pourra déterminer que cette action médicale était intentionnelle). Ceci permettra à la docteure de prélever ses organes afin de vous faire vivre ainsi que les quatre autres patients ; (2) la docteure réalise l'opération bénigne de la nouvelle patiente dans les règles de l'art. Elle rentrera sous peu chez elle sans que ses organes n'aient été prélevés mais vous, ainsi que les quatre autres patients, mourrez.

Annexe F : Exemple d'application des modèles d'analyse conjointe multi-attributs dans le dilemme du chirurgien (*Transplant dilemma*)

Dans cet exemple, nous avons ajouté quatre attributs que nous avions énumérées dans la section présentant ce type de modèle. Les attributs qui n'interviennent pas ont été grisées car ils apparaissent comme égaux dans les deux options (i.e., certitude du résultat et dérobade) ou l'information n'est pas présente dans le dilemme (i.e., qualité des individus).

Etape 1 : La caractérisation

Attributs	**Sauver les cinq**	**Sauver l'individu isolé**
Optimisation	Oui/Non	Oui/Non
Normes morales violées	Oui/Non	Oui/Non
Impact sur l'image personnelle	+/-/0	+/-/0
Impact sur l'image sociale	+/-/0	+/-/0
Regrets	0/+/++	0/+/++
Dérobade	Oui/Non	Oui/Non
Certitude du résultat	Oui/Non	Oui/Non
Qualité des individus	+/-/0	+/-/0

Le dilemme :

Vous êtes un(e) brillant(e) chirurgien(ne) spécialiste des transplantations. Face à la pandémie mondiale, vous êtes confronté(e) à des choix que jamais auparavant vous n'avez dû faire.

Au sein de votre service se trouvent cinq patients nécessitants chacun un organe différent sans lequel il mourra. Malheureusement, aucun organe n'est disponible pour ces transplantations.

Arrive alors un(e) autre patient(e) en bonne santé pour une petite opération bénigne mais urgente. Pendant l'examen, vous constatez que tous ses organes sont compatibles avec les cinq autres patients mourants.

Ne sachant quoi faire, vous avez la possibilité de prendre une décision totalement atypique dans votre pratique.

Il y a deux options :

(1) vous réalisez une manœuvre létale pendant l'opération du nouveau patient (de la nouvelle patiente). Cette manœuvre causera donc sa mort mais vous n'en serez pas inquiété(e) (personne ne pourra déterminer que cette action médicale était intentionnelle). Ceci vous permettra de prélever ses organes afin de faire vivre vos cinq autres patients ;

(2) vous réalisez l'opération bénigne du nouveau patient (de la nouvelle patiente) dans les règles de l'art. Il(Elle) rentrera sous peu chez lui(elle) sans que ses organes n'aient été prélevés mais vos cinq autres patients mourront.

Dans ce type de dilemme (personnel), le fait de sauver les cinq patients mourants implique de commettre un acte qui viole les normes morales. De plus, le fait de réaliser une manœuvre létale intentionnellement, même si elle vise à sauver un plus grand nombre d'individus, aura un impact négatif sur l'image personnelle et sociale (même si dans l'exemple, personne ne peut déterminer que l'action était intentionnelle). On peut toutefois imaginer que l'individu puisse se poser la question : « Et si on apprenait que mon action était volontaire ? » ; auquel cas, l'impact social sera négatif dans l'option « sauver les cinq ». Enfin, pour un individu lambda, on peut supposer que le fait de réaliser intentionnellement une manœuvre létale induise inéluctablement des regrets chez le décideur.

Etape 2 : La cotation et le comptage

Attributs	**Sauver les cinq**	**#**	**Sauver l'individu isolé**	**#**
Optimisation	Oui	1	Non	2
Normes morales violées	Oui	2	Non	1
Impact sur l'image personnelle	-	2	+	1
Impact sur l'image sociale	(0)	1	(0)	1
Regrets	++	2	0/+	1
Dérobade	Non		Non	
Certitude du résultat	Oui		Oui	
Qualité des individus	0		0	
Total		8		6

L’option qui comptabilise le nombre de points le plus bas (« sauver l’individu isolé ») est à privilégier. Il s’agit donc de ne pas réaliser une manœuvre létale sur le(la) nouveau(velle) patient(e) et c’est cette option qui est choisie dans 90% des cas (Nasello et al., 2021a ; Nasello et al., 2023 ; Nasello & Triffaux, *in press*).

Annexe G : Exemple d'application des modèles d'analyse conjointe multi-attributs dans le dilemme de l'entreprise (*company's dilemma*), perspective du représentant

Comme pour notre exemple précédent, nous avons cinq attributs centraux (optimisation, normes morales violées, impact sur l'image personnelle et sociale, et regrets) et ceux qui n'interviennent pas ont été grisées.

Etape 1 : La caractérisation

Attributs	**Sauver les cinq**	**Sauver l'individu isolé**
Optimisation	Oui/Non	Oui/Non
Normes morales violées	Oui/Non	Oui/Non
Impact sur l'image personnelle	+/-/0	+/-/0
Impact sur l'image sociale	+/-/0	+/-/0
Regrets	0/+/++	0/+/++
Dérobade	Oui/Non	Oui/Non
Certitude du résultat	Oui/Non	Oui/Non
Qualité des individus	+/-/0	+/-/0

Le dilemme :

Depuis plus de 20 ans, vous êtes employée dans une petite entreprise. Vous appréciez votre travail et bénéficiez d'un bon poste ainsi qu'une bonne rémunération. Toutefois, la crise sanitaire liée à la pandémie a mis l'entreprise franchement dans le rouge et une restructuration du personnel semble imminente.

Paniqués, les six employés de votre service (dont vous) se sont mobilisés pour faire grève car vous appréhendez des licenciements sans que des indemnités de licenciement soient versées. Vous faites donc grève pour revendiquer une sécurité d'emploi et pour obtenir des indemnités de licenciement correctes en cas de restructuration.

Pour tenter de résoudre la situation, votre patron vous convoque car vous êtes la porte-parole du mouvement. Il comprend votre position et vous explique ne pas avoir énormément d'options pour arranger la situation.

Votre patron vous propose un arrangement (seuls vous et lui serez au courant de ce dernier). Cet arrangement comporte deux options à effets immédiats et sans concertation possible. La première option est de vous garantir contractuellement à vous, et vous seule, votre emploi. Vos cinq autres collègues seront alors licenciés sans indemnités de licenciement. La seconde option est de vous licencier vous, et vous seule, sans indemnités de licenciement. Vos cinq autres collègues auront alors une garantie contractuelle de leur emploi.

Comme mentionné, dans les deux cas, personne ne saura que cette décision émane de vous.

N.B. : il n'y a pas de concertation avec le patron, vous êtes la seule à décider.

Si vous choisissez la première option, vous bénéficierez d'une garantie d'emploi. Toutefois, votre employeur licenciera vos cinq autres collègues sans indemnités.

Si vous choisissez la seconde option, vous serez licenciée sans indemnités. Toutefois, votre employeur garantira l'emploi de vos cinq autres collègues.

Dans ce dilemme et cette perspective, la situation est plus complexe car, étant donné que l'individu est personnellement concerné par la décision, la manière dont vont être évaluées les différents attributs et comment ils vont être pondérés va varier significativement d'un décideur à l'autre. Par exemple, au niveau des normes morales violées, on peut considérer que le fait de sauver son emploi enfreigne les prérogatives prosociales (décision qui sera jugée égoïste). D'autres n'y verront pas une infraction aux normes morales. On peut supposer également qu'un choix égocentrique ait un impact négatif sur l'image personnelle et sociale (même si, une fois encore, personne ne peut déterminer que la décision émanait du décideur) et que cela induise des regrets chez certains ; quand d'autres penseront, par exemple, qu'en sauvegardant leur emploi, ils sauvegardent également la pérennité du niveau de vie familial.

Etape 2 : La cotation et le comptage

Attributs	**Sauver l'emploi des cinq**	**#**	**Sauver son emploi**	**#**
Optimisation	Oui	1	Non	2
Normes morales violées	Non	1	Oui	2
Impact sur l'image personnelle	+	1	-	2
Impact sur l'image sociale	(0)	1	(0)	1
Regrets	0	1	+	2
Dérobade	Non		Non	
Certitude du résultat	Oui		Oui	
Qualité des individus	0		0	
Total		5		9

Dès lors, pour un individu qui mettra l'accent sur des valeurs morales prosociales et pour qui le fait d'enfreindre ces normes aura davantage de conséquences négatives au niveau post-décisionnel (comme illustré dans le tableau ci-avant), on voit assez clairement qu'il va privilégier l'option utilitariste.