



UNIL | Université de Lausanne

Faculté des Sciences Sociales et Politiques
Institut de Psychologie sociale

Session d' Août 2012

**LES EXPLICATIONS NAÏVES DE COMPORTEMENTS :
L'EFFET DES VARIABLES DE DÉSIRABILITÉ ET D'UTILITÉ SOCIALE
SUR LA THÉORIE NAÏVE D'ACTION INTENTIONNELLE**

Mémoire de Maîtrise Universitaire en Sciences Sociales

Présenté par Anatolia Batruch

Directeur : Dompnier Benoît

Expert : Smeding Annique

Résumé

Les hypothèses traditionnelles de la théorie d'attribution causale suggèrent que les explications naïves de comportement se réfèrent soit à des causes internes, soit externes. Malle (2004) présente une approche alternative, nommée la théorie naïve d'action intentionnelle, centrée sur le concept d'intentionnalité et sur les modes d'explication qui en découlent. Dans ce travail, nous avons tenté de reproduire les hypothèses d'asymétrie acteur-observateur proposées par ce modèle, en les comparant à l'effet d'asymétrie sur les catégories provenant de la théorie traditionnelle d'attribution causale (interne/externe, présence de traits de personnalité/absence de traits de personnalité). Dans un deuxième temps, nous avons étudié l'effet de la valeur sociale sur les modes d'explication de la théorie naïve d'action intentionnelle. Plus précisément, nous avons envisagé la possibilité que l'utilisation des catégories d'explication issues de la théorie naïve d'action intentionnelle pouvait varier en fonction de deux composantes de la valeur sociale : la désirabilité sociale et l'utilité sociale (Cambon, Djouari, & Beauvois, 2006). Les résultats de cette étude indiquent que les hypothèses d'asymétrie prédites par Malle (2004) peuvent être partiellement vérifiées dans un contexte académique suisse. Dans la deuxième partie de notre étude, nous avons pu observer un effet des consignes de désirabilité et d'utilité sociale sur certaines catégories du modèle de Malle (2004). Les résultats ont également indiqué que l'utilisation de traits et d'explications internes augmentait en condition d'utilité sociale, suggérant un effet de la norme d'internalité (Beauvois & Dubois, 1988). Toutefois, l'étude de l'interaction entre les consignes de valeur sociale et la perspective acteur-observateur s'est révélée décevante.

Abstract

Traditional attribution theory hypothesis suggest explanations of behavior refer to either internal causes or external causes. Malle (2004) presents an alternative approach: the folk-conceptual theory of behavior explanations. This theory revolves around the concept of intentionality and distingues modes and types of explanations. The aim of our study was to replicate the actor-observator asymmetries predicted by this model and compare those results to the traditional actor-observer hypothesis. The second part of our study focuses on the effect social value might have on the use of categories presented by folk-conceptual theory of behavior explanations. Specifically, we considered the possibility that the use of modes and types of explanations could vary depending on two components of social value – namely social desirability and social utility (Cambon, Djouari & Beauvois, 2006). This study suggests that the actor-observator asymmetries predicted by Malle (2004) can be partially verified in a Swiss academic context. In the second part of our study, we were able to see an effect of social value on the use of several of folk theory of behavior explanations categories. Results also indicated that the use of internal explanations increased in social utility situations, suggesting an effect of the norm of internality (Beauvois & Dubois, 1988). However, the study of the interaction between social value and the actor-observer asymetry has been disappointing.

Depuis la publication de l'ouvrage d'Heider en 1958, les travaux en psychologie sociale n'ont cessé d'aborder la question de l'explication profane des événements psychologiques quotidiens (Heider, 1958; Jones & Davis, 1966; Kelley, 1967, cité par Malle, 2004). L'une des problématiques les plus étudiées sur la question est l'analyse de l'asymétrie entre les perspectives d'acteur et d'observateur, proposée notamment par Jones et Nisbett (1972) (Malle & Knobe, 1997a). L'hypothèse repose sur le postulat que les différences de perspective (acteur ou observateur) impliquent des variations dans la manière pour les acteurs et les observateurs d'expliquer un même comportement (Malle, Knobe, & Nelson, 2007).

Dans la perspective traditionnelle de la théorie d'attribution, les explications de comportement peuvent être divisées en deux facteurs : les causes situationnelles (ou externes) et les causes dispositionnelles (ou internes) (Jones & Davis, 1966; Shaver, 1975). En s'inspirant de cette conceptualisation de l'explication de comportement, l'hypothèse avancée par Jones et Nisbett (1972) suggère que les acteurs ont tendance à attribuer la cause de leurs actions à des facteurs externes, alors que les observateurs préféreraient attribuer la cause des mêmes actions à des facteurs internes. Les auteurs postulent que cette asymétrie peut être attribuée à deux facteurs : l'attention et la différence d'information. D'après eux, le comportement d'un acteur est plus saillant pour l'observateur (car il est perçu de l'extérieur), qu'il ne l'est pour les acteurs eux-même. De ce fait, les observateurs ont tendance à attribuer la cause du comportement d'une autre personne à elle-même. De plus, les observateurs ne possédant pas toutes les informations concernant le comportement, choisiraient par défaut d'attribuer

la cause du comportement à une personne, à moins de connaître d'autres facteurs extérieurs. Autrement dit les observateurs, n'ayant pas d'autres informations que la vision du comportement, auront tendance à focaliser sur le comportements et non sur les facteurs environnants (Gosling, 2007). Cette asymétrie est aujourd'hui communément acceptée en psychologie sociale (Watson, 1982; Baron, Byrne, & Branscombe, 2006; Fiske, 2009).

Néanmoins, sur la base d'une méta-analyse de 173 études, Malle (2006) a publié un article suggérant, que les résultats empiriques ne confirment pas cette hypothèse.

D'après Malle, les résultats obtenus lors de la méta-analyse sont dus à des défauts inhérents à la théorie classique de l'asymétrie acteur-observateur et à la méthodologie habituellement utilisée. Il explique que la division conceptuelle dichotomique interne/externe, sur laquelle repose la théorie de l'asymétrie acteur-observateur, est trop simpliste pour suffire à expliquer la totalité des types d'explication naïve de comportement. D'autre part, il ajoute que les études effectuées dans les laboratoires reflètent mal la réalité des explications naïves présentes au quotidien. Selon Malle, le meilleur moyen d'appréhender les explications naïves de comportements est de demander aux participants de mentionner des comportements qui les ont spontanément intéressés et non de leur soumettre des exemples de comportements à commenter (Malle, 2006).

Malle propose d'utiliser un nouveau cadre conceptuel pour tenter de reproduire le schéma d'explication causale naïf. « The Folk-Conceptual Theory of Behavior Explanation » que nous traduirons par « la théorie naïve

de l'action intentionnelle » apporte une nouvelle vision du processus d'explication de comportement.

La Théorie Naïve de l'Action Intentionnelle

La théorie naïve de l'action intentionnelle de Malle (2004) comporte trois niveaux d'analyse. Le premier niveau d'analyse précise le cadre conceptuel de l'explication naïve de comportement et explique les outils qui sont utilisés pour le représenter - principalement les différents modes et types d'explication. Le deuxième niveau explore les processus psychologiques qui sous-tendent ce cadre conceptuel. Ces processus permettent de générer, sélectionner et moduler les explications. Le troisième niveau d'analyse se situe sur le plan linguistique. Dans ce chapitre l'auteur indique les outils utilisés pour communiquer les explications et comment ceux-ci sont adaptés aux buts de l'explicateur (Malle, 2011).

Une Nouvelle Approche de l'Activité Explicative

L'ouvrage d'Heider (1958) est reconnu comme ayant été le premier à proposer la différenciation interne/externe (Kassin, Fein, & Markus, 2008). Toutefois, malgré l'impact très important de ce travail princeps, Malle (2004) suggère que les écrits d'Heider ont été faussement interprétés par Kelley (1967, cité par Malle, 2004). D'après Malle (2004), l'intention d'Heider était de séparer les causalités en facteurs personnels et impersonnels, indiquant de ce fait que les comportements intentionnels étaient à séparer des comportements non-intentionnels. La théorie naïve de l'action intentionnelle de Malle (2004) propose de remédier à cette

confusion et postule que la distinction centrale de l'explication naïve de comportements repose principalement sur le concept d'intentionnalité (Guglielmo, Monroe & Malle, 2009 ; Malle & Nelson, 2003 ; Monroe & Malle, 2010). Les personnes percevraient et conceptualiseraient différemment les actes intentionnels (par exemple : acheter une voiture) et les actes non-intentionnels (par exemple : éternuer ou tomber). Le concept d'intentionnalité serait également instinctivement compris par la population. D'après une étude réalisée par Malle et Knobe (1997a), les participants auraient une conception commune de l'intentionnalité et jugeraient habituellement les comportements intentionnels de façon similaire. Dans le cadre de l'étude, les chercheurs ont demandé aux participants de lire la description de 22 comportements et de juger du niveau d'intentionnalité du comportement sur une échelle de 1 à 8. Une explication du concept d'intentionnalité a été distribuée à la moitié de l'échantillon, mais cette variable n'aurait eu aucun effet, d'après Malle (2004). Lorsque l'on prenait le résultat d'une personne en comparaison avec le reste du groupe, la moyenne de corrélation se situait à $r(20) = .80$ (Malle & Knobe, 1997a).

Mais si l'intentionnalité apparaît comme un élément central de la théorie naïve de l'explication causale, il reste que ce concept repose sur plusieurs paramètres. En effet Malle et Knobe (1997b) proposent cinq composantes du concept de l'intentionnalité. Ainsi, pour que l'on puisse considérer un comportement comme étant intentionnel, il faut que : 1) l'acteur ait le désir d'arriver au résultat obtenu ; 2) qu'il croie que ce comportement va le mener à ce résultat ; 3) qu'il ait l'intention d'effectuer ce comportement ; 4) qu'il ait les compétences de le faire et enfin 5) qu'il

soit conscient qu'il réalise son intention en effectuant ce comportement. Au quotidien, une personne ne décomposerait pas linéairement le concept d'intentionnalité en cinq composantes, mais prendrait en compte naturellement ces facteurs pour décider de l'intentionnalité d'un acte (Malle, 2004).

Au-delà du concept d'intentionnalité, la théorie de Malle présuppose également que la population utilise qualitativement des modes distincts d'explication de comportement. Le recours à ces différents modes d'explication varierait selon que le comportement à expliquer est jugé comme intentionnel ou non (Malle et al., 2007). Ces modes d'explication ne sont pas des causes qui diffèrent par un attribut (tel que le locus de causalité ou la stabilité). Il s'agit plutôt d'approches distinctes que les gens adoptent en fonction de la nature du comportement expliqué (Malle et al., 2007). Plus particulièrement, Malle et al. (2007) proposent de distinguer les explications de cause, les explications de raison, les raisons historiques causales et les explications de facteur facilitateur. Parmi les quatre modes possibles, les deux premiers tiennent une place particulièrement importante dans le modèle proposé par Malle et al. (2007). En effet, lorsqu'un comportement est le produit d'un processus mécanique (et donc non-intentionnel), celui-ci serait expliqué par des causes. A l'inverse, une personne utiliserait plutôt des raisons pour expliquer un comportement intentionnel. Prenons un exemple proposé par Malle et al. (2007) pour illustrer la différence entre ces deux modes.

1° « Anne a étudié toute la nuit pour son examen parce qu'elle voulait le réussir. »

2° « Anne était nerveuse au sujet des résultats de son examen parce qu'elle voulait le réussir. »¹

Dans les deux cas, le comportement est expliqué par le fait qu'Anne voulait réussir son examen. Linguistiquement, l'explication est la même. Pourtant, les deux explications impliquent des inférences différentes quant aux intentions d'Anne. Dans le premier cas, le comportement de l'acteur est le fruit d'une volonté, d'un raisonnement et surtout d'une décision. Dans le deuxième cas, l'acte d'Anne n'est pas voulu. Son envie de réussir son examen est la cause de sa nervosité. Anne n'a pas décidé d'être nerveuse parce qu'elle voulait réussir (Malle et al., 2007).

Les Modes d'Explication des Comportements Intentionnels et Non Intentionnels

Afin de mieux comprendre les spécificités des différents modes d'explication proposés dans la théorie naïve d'action intentionnelle, nous allons les présenter plus en détails et ce en fonction de leur centralité dans le modèle. En effet, alors que les trois premiers modes (explication de raison, raison historique causale, explication de facteur facilitateur) concernent plus particulièrement l'explication des comportements intentionnelles, le dernier que nous évoquerons (explication de cause) s'applique aux comportements non intentionnels.

¹ (1) "Anne studied for the test all night because *she wanted to do well*." (2) "Anne was nervous about the test results because *she wanted to do well*." (Malle et al., 2007, p. 492)

Explication de Raison.²

L'explication de raison est un des quatre modes d'explications proposé par Malle (2004). Lorsqu'une personne explique un comportement à l'aide d'explications de raison, elle met en avant les raisons conscientes (les désirs ou les croyances) qui ont directement poussé l'acteur à agir intentionnellement. Pour pouvoir citer des explications de raison, il est nécessaire d'adopter la perspective de l'acteur pour comprendre les facteurs que l'agent a pris en compte pour décider d'agir. Une étude de Malle, Knobe, O'Laughlin, Pearce et Nelson (2000) a démontré que les explications de raison perdent tout leur sens lorsque l'acteur n'a pas conscience de ces raisons. Malle (2004) appelle ce phénomène « l'hypothèse de subjectivité »³. Ce terme signifie que la personne expliquant le comportement imagine que les raisons citées ont été prises en compte par l'acteur au moment de la prise de décision. L'autre caractéristique importante des explications de raison est « l'hypothèse de rationalité »⁴. La personne expliquant un comportement part du principe que l'acteur a suivi une logique rationnelle lors de sa prise de décision. Autrement dit, la personne ne pense pas seulement qu'il y a une relation causale entre les raisons évoquées et l'acte, mais aussi que ces raisons découlent d'un raisonnement logique (Malle, 2004).

Quand une personne propose une explication de raison, deux possibilités s'offrent à elle. Elle peut choisir de mentionner un désir ou une croyance, qui sont deux types de raison distincts. Les désirs marquent la

² Traduction libre de l'expression « reason explanation » (Malle, 1999b)

³ Traduction libre de l'expression « subjectivity assumption » (Malle, 2004)

⁴ Traduction libre de l'expression « rationality assumption » (Malle, 2004)

volonté de l'acteur de voir des buts se réaliser. Les croyances, en revanche, sont les connaissances que possède l'acteur, qui lui permettent de se représenter la réalité et de prendre en considération les conséquences de ses actions. Le fait de choisir un type ou l'autre n'est pas toujours psychologiquement significatif, mais peut malgré tout l'être dans certaines situations. D'une part, les croyances apportent plus de détails idiosyncrasiques sur le processus de prise de décisions. D'autre part, les croyances se réfèrent aux connaissances et aux réflexions de l'acteur, mettant en avant son côté rationnel (Malle et al., 2000).

Au niveau linguistique, il y a deux manières d'exprimer une croyance ou un désir. L'explicateur peut décider d'utiliser des « marqueurs d'état mental »⁵ pour mettre l'accent sur le type de raisons cités (Malle et al., 2000). Par exemple, l'on peut dire : « Je me suis réveillé tôt parce que *je voulais* faire du sport avant d'aller travailler. » ou « Je me suis réveillé tôt pour faire du sport avant d'aller travailler. » Certaines expressions (telles que penser, vouloir, croire ou avoir envie) servent de marqueurs d'état mental. Leur utilisation n'apporte pas d'informations supplémentaires sur le contenu de la raison, mais sur la manière de l'exprimer. Les explicateurs auront tendance à les utiliser en fonction de leurs besoins sociaux. Dans certains cas, ils le feront pour renvoyer une meilleure image d'eux-mêmes. Dans d'autres cas, il peut s'avérer utile d'employer un marqueur d'état mental pour signaler une distance entre l'explicateur et les raisons de l'acteur. Ce n'est pas la même chose de dire : « Il veut faire un régime

⁵ Traduction libre de l'expression « Mental State Markers » (Malle, 2004)

parce qu'il est gros. » que de dire : « Il veut faire un régime parce qu'il pense être gros. »

Raison Historique Causale.⁶

Dans la majorité des cas, une personne utilisera plutôt une explication de raison pour expliquer un comportement. Mais, dans certains cas, l'explicateur attribuera les causes d'un acte à des facteurs historiques causaux des explications de raison, c'est-à-dire des facteurs en arrière-plan qui ont contribué indirectement à l'acte (Malle, 2004). L'explicateur ne fait pas d'hypothèse de subjectivité ou de rationalité en invoquant une raison historique causale car l'agent n'a pas forcément conscience de l'influence de ces raisons, qui peuvent être entre autres des raisons inconscientes, des traits de personnalité, l'éducation ou des facteurs culturels. Si quelqu'un dit : « Mia travaille trop parce c'est comme ça au Japon. Tout le monde travaille trop là-bas. », on peut imaginer que Mia n'a pas forcément conscience qu'il s'agit là de la raison qui la pousse à agir ainsi. Mia n'a peut-être pas décidé d'agir ainsi, parce que « c'est comme ça au Japon », mais probablement parce qu'elle a beaucoup de travail et qu'elle veut le faire bien. L'acte est intentionnel, mais les raisons citées n'ont pas toujours été prises en considération par l'agent (Malle et al., 2007). Une raison historique causale sera plutôt utilisée si l'explicateur n'a pas accès à la raison directe du comportement ou lorsqu'il a du mal à inférer cette raison. Il est alors plus facile de tenir compte d'informations générales concernant l'acteur ou le contexte environnant.

⁶ Traduction de libre l'expression « Causal History of Reason » (Malle, 2004)

Explications de Facteur Facilitateur.⁷

La réalisation d'un acte peut dépendre également de facteurs contextuels qui ne dépendent pas des intentions de l'acteur. C'est le cas notamment lorsqu'un explicateur invoque des compétences ou des circonstances facilitatrices. Ce mode d'explication implique une intentionnalité, mais ne mentionne pas des raisons pour expliquer un comportement. En règle générale, ces explications sont utilisées pour expliquer la raison d'une performance réussie. Leur occurrence est plutôt rare (Malle, 2004).

Explication de Cause.⁸

Nous avons examiné les trois modes d'explication (les raisons, les raisons historiques causales et les facteurs facilitateurs) qui permettent d'expliquer les comportements intentionnels. Selon la théorie naïve d'action intentionnelle, les actes non-intentionnels requièrent seulement un mode explicatif : les explications de cause. Dans cette situation, l'explicateur ne fait que citer les facteurs qui ont provoqué le comportement. Les comportements non-intentionnels nécessitent peu d'inférence, ou de suppositions de la part de l'explicateur. De ce fait, les explications de cause peuvent être analysées au travers de la grille de lecture de la théorie classique d'attribution (à l'aide des catégories interne/externe, trait/non-trait; Malle et al., 2007).

⁷ Traduction libre de l'expression : « Enabling Factors Explanations » (Malle, 2004)

⁸ Traduction libre de l'expression : « Cause Explanation » (Malle, 2004)

Les Processus Psychologiques à la Base du Choix d'un Mode Explicatif

Comme nous l'avons vu, quatre modes d'explication distincts peuvent être mobilisés pour expliquer un comportement. Mais quels sont les processus psychologiques qui sous-tendent le choix d'un mode au détriment des autres? Selon la théorie naïve d'action intentionnelle, ces déterminants psychologiques peuvent être réunis en trois catégories : les jugements d'attributs de comportement, les buts pragmatiques et les ressources informationnelles⁹ (Malle, 2004).

Avant d'apporter une explication, l'explicateur va inférer certains jugements implicites sur le comportement que Malle (2004) appelle des jugements d'attributs de comportement. En premier lieu, il va porter un jugement sur l'intentionnalité de l'acte. En fonction de ce jugement, il choisira un mode d'explication adéquat. Un comportement intentionnel va provoquer l'utilisation de raisons, de raisons historiques causales ou des facteurs facilitateurs (ou plusieurs à la fois). Sinon, le choix va directement se porter sur l'explication de cause. L'explicateur prend également en compte un deuxième attribut important - la difficulté de la tâche. Si la tâche est jugée difficile, l'explicateur choisira en priorité les facteurs facilitateurs. Finalement, un comportement va être considéré différemment selon qu'il semble être un cas particulier unique ou s'il s'inscrit dans une tendance. L'occurrence de raisons historiques causales augmente considérablement en présence de tendances.

Les buts pragmatiques incluent tous les projets que l'explicateur cherche à accomplir, en expliquant un comportement. L'explication de

⁹ Traduction libre des expressions : « Judged behavior attributes », « Pragmatic goals » et « Information resources » (Malle, 2004)

comportement peut remplir de nombreuses fonctions, telles que la dissipation d'une confusion, la gestion d'une image personnelle lors d'une interaction, ou même la décharge de responsabilité. Il y a deux types de buts qui varient en fonction de la personne qui en bénéficie. Certains buts sont tournés vers l'interlocuteur. Malle appelle cette catégorie « l'adaptation au public »¹⁰. Les explicateurs sont influencés par les attentes d'un interlocuteur. En cherchant à apporter les informations pertinentes, ils adaptent leur explication (et le mode d'explication) en fonction des questions posées par l'interlocuteur. D'autres buts profitent à l'explicateur. Ce sont les « gestions d'impression »¹¹. La gestion d'impression est un acte d'influence sociale, qui va conditionner le choix du mode d'explication. Une personne souhaitant paraître rationnel, par exemple, aura tendance à utiliser plus de croyances que de désirs. Il est également plus fréquent de rencontrer des raisons historiques causales lorsqu'un explicateur fait référence à des comportements négatifs (Malle et al., 2000). Bien que ces adaptations soient plus visibles lors d'interactions, elles sont aussi présentes lors d'explications privées (Malle, 2004) car un explicateur peut accommoder ses explications pour justifier ses comportements à ses propres yeux (Steele, 1988).

Enfin, le mode d'explication dépend également des ressources informationnelles à disposition. Un manque d'information peut forcer un explicateur à ajuster ses buts ou ses explications. Plusieurs études menées par Malle (2004) ont démontré que si l'on manipule la variable de l'accès à l'information, le mode d'explication s'adapte également. Par exemple, plus un observateur a d'informations sur l'action et l'agent, plus son explication

¹⁰ Traduction libre de l'expression : « Audience design » (Malle, 2004)

¹¹ Traduction libre de l'expression : « Impression management » (Malle, 2004)

se rapproche de celui de l'acteur. L'asymétrie entre acteur et observateur peut donc diminuer en fonction des connaissances de l'explicateur. L'explicateur s'aide souvent d'inférences pour comprendre un comportement inconnu. Ceci a pour effet d'augmenter le nombre de raisons historiques causales, puisque l'explicateur est contraint de puiser des explications de règles générales.

La Théorie Naïve d'Action Intentionnel : Intégration et Synthèse

En résumé, la théorie naïve d'action intentionnelle (Malle, 2004) postule que les explications naïves de comportements sont, tout d'abord, divisées en deux catégories : les comportements intentionnels et non-intentionnels. Si le comportement est jugé comme intentionnel par l'explicateur, il peut choisir entre trois modes d'explications : les explications de raison, les raisons historiques causales et les facteurs facilitateurs. A priori, si l'explicateur est bien informé, il se dirigera spontanément vers les raisons. Au sein de la catégorie des explications de raison, on trouve deux types d'explications : les croyances et les désirs. Un explicateur choisit de les expliquer avec ou sans marqueur d'état mental. Toutefois, certains processus psychologiques peuvent interférer dans le choix du mode et pousser un explicateur à choisir les raisons historiques causales. Enfin, si le comportement fait référence à une performance, alors il est possible que l'explicateur mentionne des facteurs facilitateurs.

Bien entendu, il arrive fréquemment qu'un explicateur utilise une combinaison de mode lors de son explication, et ce d'autant plus s'il perçoit un sentiment de confusion chez son interlocuteur. Si le comportement n'est

pas intentionnel, alors l'explicateur citera des facteurs causaux mécaniques, c'est-à-dire des explications de cause. (Figure 1)

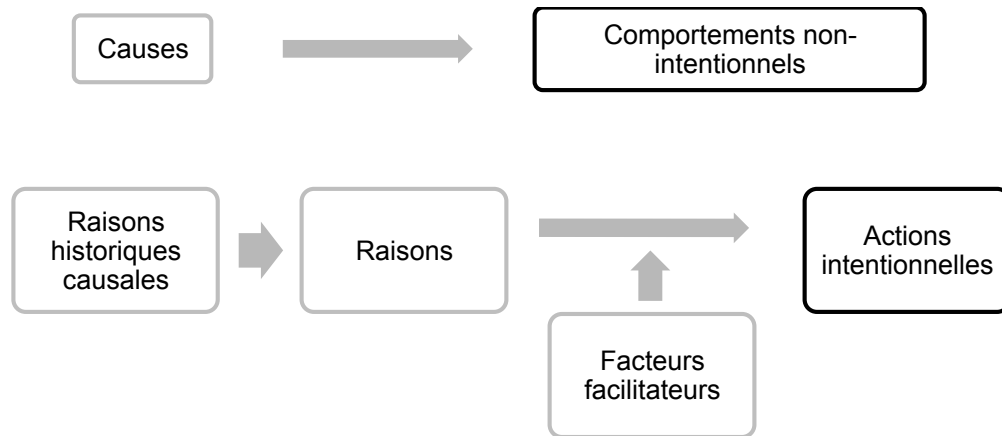


Figure 1. Quatre modes d'explication pour les comportements intentionnels et non-intentionnels (Malle 2003).

A l'aide du modèle de la théorie naïve d'action intentionnelle, Malle (2004) prédit quatre asymétries acteur-observateur, qui peuvent être partiellement expliquées par des processus psychologiques. La première hypothèse se rapporte à l'intentionnalité du comportement cité. D'après Malle et Knobe (1997a), les acteurs mentionnent des comportements moins intentionnels que les observateurs et ce pour deux raisons. D'une part, les personnes choisissent en général des comportements qui les ont interpellés. Les acteurs, connaissant en général les raisons de leurs actions intentionnelles, préfèrent proposer des comportements moins intentionnels. D'autre part, pour un observateur, les comportements intentionnels apportent plus d'informations sur l'agent. Il est plus facile de tirer des conclusions d'un comportement intentionnel. Il est donc logique, d'après lui, que les observateurs soient attirés par des comportements intentionnels.

La deuxième hypothèse postule que les acteurs utilisent plus d'explications de raison et moins de raisons historiques causales que les observateurs. Les acteurs ont un accès cognitif direct aux raisons qui l'ont poussé à agir. Cet accès pourrait les inciter à choisir des raisons qu'ils connaissent puisqu'ils les avaient eues en tête au moment d'agir. A l'inverse, un observateur n'a pas forcément eu accès au processus de décision. Il doit donc baser ses explications sur des inférences sur la personnalité de l'acteur, sur l'action et sur le contexte. Ces inférences augmentent la probabilité d'invoquer des règles générales lors d'explications, et donc d'utiliser des raisons historiques causales. Il se peut également que la perspective de l'acteur soit influencée par des buts de gestion d'impression. Knobe et Malle (2002) ont démontré que les acteurs utilisant des explications de raison étaient perçus comme des agents conscients, rationnels avec un libre choix, alors que les raisons historiques causales mettaient en avant des facteurs hors du contrôle de l'acteur (Malle et al., 2000). Si l'on part du principe que paraître rationnel est socialement valorisé dans notre société, on peut présupposer que l'acteur veuille renvoyer cette image de lui-même et préférera les explications de raisons. Il est difficile d'imaginer, par contre, qu'un observateur bénéficie du fait de renvoyer une image rationnelle de l'agent. Cette hypothèse pourrait donc être expliquée par deux des processus psychologiques cités par Malle (2004) : les ressources informationnelles et la gestion d'impression.

La troisième hypothèse concerne un type de raison, ici les croyances. Selon cette hypothèse, les acteurs préféreraient citer plus de croyances et moins de désirs que les observateurs. Malle et al. (2007) pensent que l'on

puisse attribuer cette asymétrie à une différence des ressources informationnelles. Les désirs et les croyances sont tous deux des raisons. Nous pouvons donc imaginer que la mémoire de l'acteur ne soit pas un facteur déterminant de ce processus, puisque rien ne porte à croire que l'acteur se souviendrait mieux d'une catégorie de raison, que d'une autre. Dans ce cas, il semble plus plausible que ce soit la perspective d'observateur qui ait un impact sur cette hypothèse. Malle et al. (2007) suggèrent que l'observateur a plus de mal à inférer des croyances que des désirs. Tout d'abord, un désir peut être visible. Prenons l'exemple d'une personne en train de manger. Un observateur dirait probablement qu'elle avait envie de manger, plutôt que de dire, par exemple, que la personne mangeait parce qu'elle pensait être occupée plus tard. Il est plus difficile d'inférer des croyances car ce sont des représentations idiosyncratiques, caractéristiques de la personne et de son environnement (Malle et al., 2007). Selon une autre étude (Malle et al., 2000), il apparaîtrait que les croyances projeteraient également une impression de rationalité. Il est donc possible qu'un acteur mentionne des croyances pour cette raison-là.

La quatrième hypothèse suggère que les acteurs utilisent plus fréquemment des marqueurs d'état mental que les observateurs, lorsqu'ils citent des croyances. Malle (2004) suggère qu'il n'est pas toujours nécessaire linguistiquement d'employer des marqueurs d'état mental pour un acteur. Une personne qui dit : « J'ai pris un parapluie parce qu'il pleut. » veut probablement dire la même chose que la personne qui dit : « J'ai pris un parapluie parce que je crois qu'il pleut. ». Il n'y a donc pas réellement de valeur ajoutée au fait d'utiliser un marqueur d'état mental pour un acteur.

Un acteur représente directement le contenu de sa croyance lors de son explication. Ceci peut expliquer le peu d'occurrence de croyances marquées mentionnées par l'acteur. Inversement, un observateur tente d'exprimer le contenu de la croyance d'une autre personne. Il est donc plus important pour lui de préciser qu'il s'agit d'une croyance. Les marqueurs d'état mental servent aussi à transmettre un jugement sur la croyance citée. Ces marqueurs installent une distance entre l'observateur et les raisons qu'ils mentionnent. Si nous reprenons l'exemple du parapluie, l'observateur qui choisit de dire : « Il a pris un parapluie parce qu'il croit qu'il pleut. » émet peut-être des doutes sur la croyance de l'agent. L'utilisation du marqueur permet de montrer qu'il ne partage pas forcément les croyances de l'agent (Malle, 2004).

Nous avons pu voir que, selon la théorie naïve d'action intentionnelle, la volonté de renvoyer une bonne image de soi-même peut influencer la manière qu'une personne a de décrire un comportement. D'après Malle (2004), le choix du mode d'explication peut considérablement varier en fonction des buts sociaux de l'explicateur. Toutefois, force est de constater que cette question n'a été que peu traitée dans le cadre de la théorie naïve d'action intentionnelle. Quel serait l'effet d'une manipulation de la variable de gestion d'impression sur le choix de mode et de type d'explications ? Sur ce point, les recherches menées dans d'autres cadres théoriques ayant étudié les conséquences des stratégies d'autoprésentation sur l'activité explicative peuvent apporter des éléments nouveaux et ainsi enrichir l'approche théorique développée par Malle (2004). C'est notamment le cas des travaux conduits depuis plus de 30 sur la

théorie de la norme d'internalité (Beauvois & Dubois, 1988; Dubois & Beauvois, 2008; Jellison & Green, 1981).

La Théorie de la Norme d'Internalité

Les premières recherches ayant cherché à mettre en évidence l'impact des normes sociales sur l'activité explicative remontent aux travaux princeps conduits par Jellison et Greene (1981) qui ont proposé le concept de norme d'internalité. Dans leur article, ces auteurs suggéraient que les gens bénéficieraient d'une valorisation sociale si ceux-ci mettaient en valeur des explications internes. Beauvois et Dubois (1988) sont partis de ce constat pour approfondir la thématique de la valorisation sociale des explications internes. La norme d'internalité serait, d'après eux, une norme sociale générale valorisant le poids causal de l'acteur dans les explications naïves de comportements. La norme d'internalité suggérerait aux gens de proposer des explications internes quelle que ce soit la valence (positive ou négative) du comportement expliqué. L'explication interne est donc toujours valorisée même lorsque que l'événement évoqué est négatif (Beauvois & Dubois, 1988).

Une Approche Sociétale de l'Activité Explicative

La norme d'internalité n'est pas un processus psychologique propre à la personne qui explique le comportement, mais plutôt une nécessité sociale. Beauvois et Dubois (2008) définissent une norme sociale comme « une attribution de valeur sociale à un type d'événement inscrit dans une structure sociale. Une valeur sociale est définie comme une conscience des

gens d'une nécessité spécifique à cette structure sociale. », (p. 1740). La norme d'internalité est une valeur sociale ancrée dans un fonctionnement sociétal individualiste et libéral (Beauvois & Dubois, 2009). La norme d'internalité n'est pas une norme descriptive, destinée à décrire ce que les gens font. C'est une norme prescriptive indiquant aux gens ce qu'ils devraient faire dans un contexte d'évaluation pour bénéficier d'une valorisation sociale (Dubois & Beauvois, 2008).

Encouragés par la norme d'internalité, il apparaîtrait que les individus attribuent davantage d'explications internes lorsqu'ils interagissent avec une personne à laquelle ils accordent une valeur (Dubois, 1994 ; Gilibert et Cambon, 2003). Ces résultats apparaissent lorsque le participant est invité à répondre à la place d'une personne possédant une certaine valeur sociale (paradigme d'identification). Par exemple, on demande au participant tour à tour de répondre à la place d'un bon ou mauvais élève, d'un ami ou ennemi, etc. Il semblerait aussi que les participants attribuent plus d'explications internes en situation d'évaluation, face à un enseignant, par exemple, ou une personne susceptible de donner un emploi (paradigme d'autoprésentation). Finalement, l'inverse s'avère être aussi vrai. D'après Cambon, Djouari et Beauvois (2006), lorsqu'on est soi-même juge des explications d'une personne, nous valorisons plus les explications internes qu'externes (paradigme des juges). Ces résultats indiqueraient, selon Cambon et al. (2006), l'existence d'une valeur sociale attribuée aux explications internes, indépendamment des différences individuelles.

La Désirabilité Sociale et l'Utilité Sociale: Deux Composantes de la Valeur Sociale

L'utilisation d'explications internes n'est pas le seul outil à disposition des gens pour s'autoprésenter positivement. Les explications personnologiques (l'utilisation de traits de personnalité) pourraient également être valorisées socialement dans des contextes évaluatifs (Dompnier & Pansu, 2010; Dubois & Aubert, 2010). Contrairement à la vision traditionnelle qui considère que les traits de personnalité sont des outils descriptifs, Beauvois et Dubois (2009) soutiennent que les traits sont des critères évaluatifs. Ils indiquent la valeur sociale d'une personne. Les traits informent du type de comportements qu'on peut attendre d'une personne, mais aussi comment d'autres peuvent se comporter avec cette personne (en fonction du trait cité). Beauvois et Dubois (2009) postulent qu'il y a deux aspects à la valeur sociale : la désirabilité et l'utilité sociale.

La désirabilité sociale correspondrait à une valeur chargée affectivement, attribuée à une personne en fonction de sa relation aux autres. Elle met en opposition des traits considérés comme positifs ou négatifs relatifs à la sociabilité d'une personne (Cambon et al., 2006). Selon Cambon et al. (2006), l'utilité sociale « reflète les connaissances que nous possédons sur la probabilité de succès ou d'échec d'une personne dans la vie sociale et est basée sur la capacité de la personne à remplir les exigences de la société dans laquelle elle vit. »¹² (p.168)

Selon Cambon et al. (2006), il apparaîtrait que la norme d'internalité affecte les deux aspects de la valeur sociale différemment. En effet, les

¹² Traduction libre de « This reflects the knowledge we have of a person's chances of success or failure in social life, and is based on how well that person meets the requirements of a given society in which he or she lives. » (Cambon et al., 2006, p.168)

personnes les plus affectées par la norme d'internalité (utilisant le plus d'explications internes) se décrivent comme étant des personnes plus « utiles » que « désirables ». Une autre recherche réalisée par Cambon, Djouari et Beauvois (2001, cités par Dompnier, 2006) démontre que des cibles très internes sont perçues comme moins désirables mais plus utiles que des cibles moins internes. Si l'on s'en tient à ces résultats, il semblerait que la norme d'internalité soit plus liée à l'utilité sociale qu'elle ne le serait à la désirabilité sociale (Dubois & Aubert, 2010).

A partir de ces résultats, on peut envisager qu'une différenciation des stratégies d'autoprésentation en désirabilité sociale et utilité sociale pourrait également impliquer des variations des types d'explication. Les résultats de Cambon et al. (2006) nous poussent à croire qu'en condition d'utilité sociale, les explicateurs préféreraient utiliser davantage de traits et d'explications internes qu'en condition standard ou de désirabilité sociale.

Problématique

Dans notre étude, nous utiliserons ici l'approche bidimensionnelle de la valeur sociale proposée par Cambon et al. (2006) pour voir quel serait l'effet de ces paramètres sur les hypothèses de la théorie naïve d'action intentionnelle. En effet, Malle présuppose que la gestion d'impression influe sur la manière d'expliquer un comportement. La gestion d'impression n'étant que des buts sociaux que peut poursuivre l'explicateur, nous proposons de considérer que les buts sociaux des explicateurs sont eux-mêmes influencés par une volonté de s'autoprésenter positivement. Il s'agira de voir si les deux concepts de valeur sociale (la désirabilité et l'utilité

sociale) peuvent être interprétés comme des buts distincts de gestion d'impression et si ceux-ci produiront un effet sur les explications de comportement. Nous essayerons donc de déterminer si la grille de lecture des explications de comportement développée par Malle peut être interprétée à la lumière des concepts de valeur sociale proposés par Beauvois et Dubois (2009). Il nous semble de plus intéressant de voir si, inversement, les catégories de la théorie naïve d'action intentionnelle peuvent apporter un nouvel éclairage sur l'effet de la valeur sociale présente dans les explications de comportements.

Nous essayerons tout d'abord de reproduire les hypothèses de Malle et al. (2007) dans le cadre d'une production explicative spontanée, c'est-à-dire sans consigne spécifique d'autoprésentation. Nous tenterons de voir s'il est possible de répéter les mêmes résultats dans un contexte académique suisse. Nous voulions tout de même insister sur l'utilisation des catégories habituelles de la théorie traditionnelle d'attribution causale : interne/externe, traits/non-traits. Si l'on s'en tient à la théorie de la norme d'internalité de Beauvois et Dubois (1988), nous nous attendons à ce que, dans un contexte évaluatif, les personnes utilisent plus d'explications internes ou des explications psychologiques si elles veulent renvoyer une image positive d'elles-mêmes, en particulier lorsqu'il s'agit de mettre en avant son utilité sociale. Il s'agira de voir si dans notre étude, les participants ont conscience de cette norme et s'ils l'appliquent lors de production explicative. Finalement, nous chercherons à voir si l'incitation à adopter un but précis de gestion d'impression (tour à tour la désirabilité et l'utilité sociale) produira un effet sur les prédictions de la théorie naïve d'action intentionnelle.

Méthode

Participants

123 étudiants de première année de bachelor de psychologie ont été interrogés durant le semestre d'automne 2011. Les participants provenaient des universités de Genève et de Lausanne. L'échantillon était composé de 101 femmes (82.11 %) et 22 hommes (17.89%). L'âge moyen des participants était de 21.95 ans (SD= 4.46).

Matériel et Procédure

Le questionnaire a été créé sur la base de questions utilisées dans l'étude 1 de Malle et al (2007). Il a été administré dans une classe de première année de psychologie au début d'un cours de Psychologie de développement cognitif. L'étude leur a été présentée comme une étude sur l'explication de comportement. Les participants ont répondu au questionnaire de façon individuelle. Les étudiants ont disposé de 20 minutes pour terminer le questionnaire.

Le questionnaire contenait quatre questions ouvertes. Les participants devaient se rappeler de la dernière fois qu'ils avaient essayé d'expliquer un comportement. Après avoir décrit le comportement, il leur était demandé d'apporter une explication pour le comportement. Ils étaient ensuite invités à indiquer le degré d'intentionnalité du comportement décrit sur une échelle de 1 (pas du tout intentionnel) à 7 (tout à fait intentionnel). Les questionnaires étaient séparés en deux groupes : perspective d'acteur (N = 56) ou d'observateur (N = 67). Les participants répondaient ensuite aux mêmes questions selon deux consignes spécifiques. L'une de celles-ci était

une consigne de désirabilité sociale invitant les participants à expliquer le comportement qu'ils avaient mentionné plus tôt de façon à se faire apprécier de l'enseignant. L'autre consigne était une consigne d'utilité sociale les invitant à expliquer ce même comportement de façon à montrer à leur enseignant qu'ils avaient tout ce qu'il fallait pour réussir à l'Université. L'ordre de ces consignes était contrebalancé entre les participants.

Codage

Chaque participant a décrit un comportement (total=123) et a proposé trois explications (condition standard, condition de désirabilité sociale, condition d'utilité sociale) (total=369). Toutes les explications ont été codées selon une version simplifiée du schéma de codage proposé par Malle (2004). Les catégories de codage comprennent les trois paramètres de la théorie naïve d'explication de comportement (causes/ raisons/ raisons historiques causales, désir vs croyance, marque vs absence de marque) ainsi que d'autres paramètres issus de la théorie classique d'attribution causale (explication interne vs externe, trait vs absence de trait).

Le schéma de codage de Malle prévoit pour chaque mode d'explication des sous-catégories. Nous avons décidé de choisir celles qui nous semblaient les plus pertinentes pour notre étude. Les causes, les raisons historiques causales et les facteurs facilitateurs ont été subdivisés en trois catégories : Agent, Situation et Autres. La catégorie « Agent » renvoie aux comportements provoqués par l'agent. Les comportements occasionnés par des facteurs situationnels étaient ensuite classés dans la case « Situation » du mode d'explications. La catégorie « Autres » était attribuée aux

comportements causés par une tierce personne. Chaque explication de raison a été séparée en croyance ou en désir, et celles-ci pouvaient être accompagnées ou non d'un marqueur.

Le comportement était soit jugé comme non-intentionnel par l'explicateur et allait dans une sous-catégorie des explications de cause, soit le comportement était considéré intentionnel et était ensuite placé dans une des sous-catégories (agent/situation/autres ou croyance marquée/croyance non-marquée, désir marqué/ désir non-marqué) du mode d'explications utilisé (raisons historiques causales, explication de raison ou facteurs facilitateur) Chaque comportement étaient aussi classé selon son appartenance aux catégories traditionnelles : interne/externe, trait/non-trait.

Par la suite, il s'est avéré que le nombre d'occurrences par sous-catégorie était trop petit, nous sommes donc revenus aux catégories plus générales : causes, raisons historiques causales, explication de raison. Les facteurs facilitateurs ont du être abandonné pour les mêmes raisons. Nous avons tout de même gardé les désirs, les croyances et les marqueurs d'états mentaux.

Données Manquantes

Tous les participants ont répondu aux questions de la condition standard. Toutefois, les questions des conditions de désirabilité et utilité sociale ont posé quelques problèmes. Il semblerait que les participants n'aient pas tous compris le sens des questions. Quelques réponses ne pouvaient pas être codées. Certains participants ont répondu à la question d'utilité sociale comme si celle-ci était dissociée du comportement

mentionné au départ. Dans certains cas, les participants avaient des difficultés à expliquer le comportement d'une tierce personne et ce faisant de mettre en avant ses propres qualités. Un certain nombre de personnes ne répondaient pas aux questions en utilisant des phrases complètes. Finalement, il est arrivé que les participants répondent « Idem » dans certaines conditions. Le processus de codage a été dans l'ensemble laborieux. Nous avons toutefois essayé de garder en tête le sens profond des explications pour les placer dans la catégorie la plus adéquate. Nous sommes arrivés à la conclusion que les conditions d'utilité et de désirabilité sociale n'étaient pas toujours adaptées aux explications d'observateurs. Alors que nous avons pu comptabiliser 122 explications en condition standard, nous disposions de 96 réponses codables en condition de désirabilité et 84 en condition d'utilité. Ces malentendus ont eu un impact sur le nombre de réponses des observateurs dans les deux conditions mentionnées, mais ont peut-être également affecté le contenu de ces réponses.

Résultats

Les données recueillies ont fait l'objet d'analyses spécifiques en lien avec nos différentes hypothèses.

Test des Hypothèses du Modèle de Malle (2004)

Dans un premier temps, les données permettant de tester les hypothèses du modèle de Malle (2004) ont été analysées par analyses de

variance. Ces analyses ont comparé les réponses des participants répartis entre les conditions « acteur » et « observateur ».

Concernant les productions explicatives des participants, le tableau 1 résume les pourcentages d'utilisation d'un mode précis d'explication selon la perspective de l'explicateur et dans la condition standard. Le pourcentage est basé sur la totalité des explications des acteurs ou des observateurs. Il est important de se rappeler que le total des explications n'est pas forcément 100%, car il arrive que les explicateurs utilisent plus d'une raison.

Tableau 1. Pourcentage de la probabilité d'emploi d'un mode d'explication selon la perspective adoptée dans la condition standard.

Mode / Type d'explication	Perspective d'acteur	Perspective d'observateur
Explication de cause	44%	21%
Raison historique causale	20%	55%
Explication de raison	39%	38%
Croyance (Raison)	9%	8%
Désir (Raison)	30%	30%
Présence de marqueur d'état mental	7%	18%
Absence de marqueur d'état mental	32%	20%
Explication Externe	60%	40%
Présence de traits de personnalité	12%	14%

Tout d'abord, il apparaît que la moyenne de l'intentionnalité des comportements était différente selon la perspective. En moyenne, l'intentionnalité des explications était jugée à 4.185 chez les acteurs, et à 4.851 chez les observateurs, $F(1, 120) = 3.247, p < 0.05$. Ces résultats confirment la première hypothèse du modèle. En effet, selon Malle et Knobe (1997a), les acteurs ont tendance à choisir des comportements moins intentionnels que les observateurs.

Les résultats révèlent que les occurrences des explications de cause diffèrent considérablement selon la perspective de l'explicateur, $F(1, 120) = 8.025, p < 0.01$. Parmi les explications d'acteur, 44% sont des causes ; alors que chez les observateurs, seuls 21% sont des causes. Ces résultats ne semblent pas surprenants, compte tenu du fait que le concept d'explication de cause est directement lié à l'intentionnalité d'un acte. Une cause peut seulement être produite par un comportement non-intentionnel.

L'une des hypothèses de Malle (2004) soutenait que les acteurs employaient plus d'explications de raison que les observateurs. Or les résultats indiquent que les acteurs et les observateurs emploient le même pourcentage de raisons, $F(1, 120) = 0.25, ns$. Néanmoins, nos résultats confirment l'asymétrie entre acteur et observateur pour les raisons historiques causales, $F(1, 120) = 17.572, p < 0.01$. Les observateurs emploient plus de raisons historiques causales que les acteurs. Etant donné que les observateurs utilisent moins d'explications de cause que les acteurs, ils compensent la différence avec un nombre supérieur de raisons historiques causales, confirmant partiellement l'hypothèse de Malle et al. (2007).

La troisième hypothèse concernait les types d'explication de raison. Selon Malle et al. (2007), les acteurs citeraient plus de croyances et moins de désirs que les observateurs. Notre étude présente des résultats différents. Nous n'avons pas trouvé d'asymétrie significative entre les acteurs et les observateurs que ce soit pour les désirs $F(1,120) = 0.035, ns$, ou les croyances, $F(1,120) = 0.073, ns$. En moyenne, les deux groupes ont utilisé plus de désirs (30%) que de croyances. (9%)

La dernière hypothèse postulait qu'un observateur avait besoin de plus de marqueurs d'état mental lorsqu'il mentionnait des croyances pour marquer une distance entre lui et l'agent. (Malle et al., 2007). Il est difficile de confirmer ou d'informer ce postulat, sur la base de nos résultats. Nous avons recensé seulement deux croyances marquées. Les deux étaient formulées par des observateurs, mais la taille de l'échantillon ne permet pas de généraliser ces effets. En général, nous avons trouvé une différence tendancielle d'utilisation de marqueurs (dans les deux catégories : désir et croyance), $F(1, 120) = 3.275, p < 0.1$. Les observateurs emploient plus souvent des marqueurs d'état mental que les acteurs (18% vs. 7%). Bien que ce résultat ne soit que tendanciellement significatif, il rejoint les prédictions de Malle et al. (2007) selon lesquelles les marqueurs d'état mental seraient peut-être plus utiles aux observateurs qu'aux acteurs.

Afin de resituer ces données dans le cadre de la théorie classique d'attribution causale, les critères de classification interne/externe et présence de trait/absence de trait ont également été testés. Contrairement aux résultats de la méta-analyse de Malle (2006), nous avons trouvé un effet significatif de l'asymétrie acteur-observateur sur la variable Interne/Externe, $F(1, 120) = 4.62, p < 0.05$. 41% des explications de comportement des acteurs renvoyaient à des explications externes, alors que seul 23% des explications de comportement de observateurs étaient causées par des facteurs externes. Ce résultat confirme l'asymétrie postulée par Jones et Nisbett (1972). En moyenne, les deux groupes utilisent plus d'explications internes qu'externes, cependant les acteurs présentent beaucoup plus d'explications externes que les observateurs. Enfin, La variable des traits n'a pas été

influencée par un effet d'asymétrie en condition standard, $F(1, 120) = 0.034$, ns. En moyenne, les acteurs et les observateurs ont mentionnés des traits dans les mêmes proportions.

L'impact des Consignes d'Autoprésentation sur la Production Explicative

Les données recueillies ont été ensuite soumises à de nouvelles analyses de façon à rendre compte de l'effet des consignes de désirabilité sociale et d'utilité sociale sur les productions explicatives en comparaison avec la consigne standard. Les données obtenues en consigne standard ont donc été intégrées à celles observées sous les consignes de désirabilité sociale et d'utilité sociale. Des analyses de variance de type 2 (perspective adoptée : acteur vs. observateur) x 3 (types de consignes : standard, désirabilité sociale, utilité sociale) ont été conduites sur chacun des modes d'explication. Le test de sphéricité de Mauchly's étant statistiquement significatif ($p < 0.01$) dans tous nos résultats, nous avons, par conséquent, interprété les résultats du F -test de Greenhouse-Geisser.

Le tableau 2 résume les pourcentages d'occurrence d'un mode d'explication sur la totalité des explications selon la perspective de l'explicateur et en fonction de la consigne de réponse donnée. Les moyennes en condition standard ne correspondent pas exactement à celles citées dans le tableau 1. Ceci est dû aux données manquantes dans les conditions de désirabilité sociale et utilité sociale. Les moyennes, ici, ont été comptabilisées sur la totalité des réponses valables dans les trois conditions.

L'échantillon contient 40 explications d'acteurs (48%) et 44 explications d'observateurs (52%).

Tableau 2. Pourcentage d'emploi d'un mode d'explication selon la perspective adoptée et en fonction des conditions d'autoprésentation.

Mode et type d'explications selon la perspective	Condition standard	Condition de désirabilité sociale	Condition d'utilité sociale
Explication de cause			
Acteur	43%	43%	38%
Observateur	21%	21%	23%
Raison historique causale			
Acteur	20%	25%	45%
Observateur	48%	60%	52%
Explication de raison			
Acteur	40%	30%	17%
Observateur	41%	27%	25%
Croyance (Raison)			
Acteur	13%	9%	8%
Observateur	11%	0%	0%
Désir (Raison)			
Acteur	28%	18%	8%
Observateur	27%	25%	25%
Présence de marqueur d'état mental			
Acteur	5%	8%	5%
Observateur	18%	16%	13%
Absence de marqueur d'état mental			
Acteur	35%	23%	13%
Observateur	23%	11%	11%
Explication externe			
Acteur	43%	32%	25%
Observateur	23%	30%	25%
Présence de trait			
Acteur	13%	20%	43%
Observateur	18%	13%	16%

Explication de Cause.

Les explications de cause varient peu entre les conditions. Il n'y a pas de différence significative, tant entre les consignes, $F(1.545, 126.967) = 1.44$; ns, qu'en ce qui concerne l'interaction, $F(1.545, 126.967) = 1.44$; ns.

Raison Historique causale.

Les résultats des raisons historiques causales paraissent intéressants. L'effet principal du type de consignes, $F(1.585, 129.963) = 5.260$, $p < 0.01$, et l'interaction entre la perspective adoptée et le type de consignes, $F(1.585, 129.963) = 4.714$, $p < 0.01$, sont tous deux significatifs. L'effet du type de consignes est particulièrement saillant chez les acteurs. Plus l'acteur se situe dans un contexte évaluatif teinté d'utilité, plus il utilise des raisons historiques causales. Dans la condition standard, 20% des explications des acteurs sont des raisons causales historiques. Le pourcentage augmente à 25% des explications d'acteur en condition de désirabilité, pour finir à 45% en condition d'utilité. Les observateurs, en moyenne utilisent plus de raisons historiques causales que les acteurs, résultat prédit par Malle et al. (2007). Bien que les différences chez les observateurs soient moins flagrantes entre les conditions, il paraît tout de même surprenant que les observateurs utilisent plus de raisons historiques causales dans la condition de désirabilité sociale que dans les deux autres.

La figure 2 illustre les variations des pourcentages des acteurs et des observateurs en fonction des consignes.

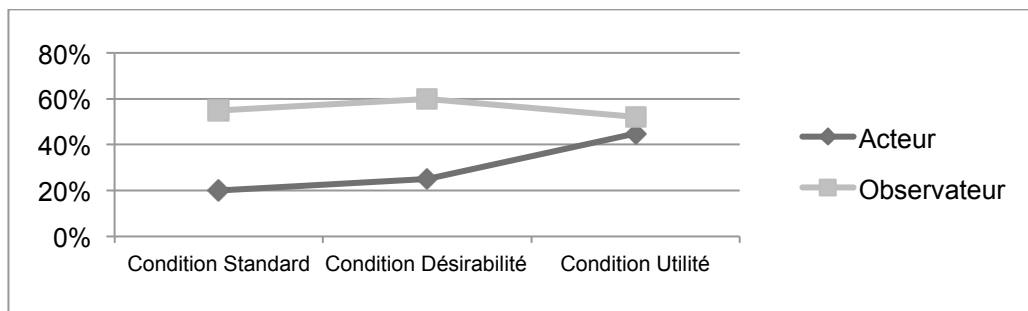


Figure 2. Moyennes des raisons historiques causales des acteurs et observateurs selon la consigne adoptée.

Explication de Raison.

Nous remarquons un effet du type de consignes dans l'utilisation d'explication de raison, $F(1.450, 118.939) = 9.923, p < 0.01$. Cet effet est probablement dû à l'augmentation d'occurrences des raisons historiques causales dans les conditions de désirabilité et d'utilité sociale. Nous pouvons observer que le pourcentage d'explication de raison des acteurs diminue drastiquement dans la condition d'utilité. En moyenne, 40% des explications d'acteurs sont des raisons dans la condition standard ; 30% dans la condition de désirabilité ; et le pourcentage descend à 17% dans la condition d'utilité sociale. Il se peut que les acteurs préfèrent employer moins de raisons, au profit de raisons historiques causales lorsqu'ils sont dans un contexte évaluatif. L'interaction avec l'asymétrie ne produit pas de résultat significatif, $F(1.450, 118.939) = 0.71 ; ns.$ Les observateurs utilisent également moins de raisons selon la condition, mais leur pourcentage varie moins. Les observateurs mentionnent 41% dans la condition standard. Le pourcentage baisse à 27% dans la condition de désirabilité sociale et réside à 25% dans la condition d'utilité sociale.

Pour les catégories de raisons (croyance, désir, présence ou absence de marqueur), nous trouvons quelques résultats significatifs. Les désirs, $F(1.553, 127.323) = 4.066, p < 0.05$, et les croyances, $F(1.207, 98.969) = 3.799, p < 0.05$, suggèrent qu'il existe un effet du type de consignes, mais peu d'effet de la perspective adoptée, $F(1.553, 127.323) = 2.581, ns$, et $F(1.207, 98.969) = 2.481, ns$. Ces résultats ne nous semblent pas particulièrement informatifs. Puisque les raisons diminuent en fonction des consignes, les désirs et les croyances vont obligatoirement diminuer aussi.

Le pourcentage d'absence de marqueur varie en fonction du type de consigne, $F(1.648, 135.163) = 10.466, p < 0.01$, mais ne varie pas significativement selon la perspective de l'explicateur, $F(1.648, 135.163) = 1.298, ns$. Plus les participants cherchent à mettre en avant leur utilité sociale, plus ils marquent leurs désirs et croyances. Cet effet pourrait indiquer une volonté pour l'observateur de se détacher de la position de l'agent. Pour l'acteur, Malle (2004) suggère que l'utilisation de marqueurs peut donner une impression de rationalité. Toutefois, il ne serait pas prudent de tirer des conclusions sur la base de ce résultat. Le nombre d'occurrence de marqueur est relativement petit.

Catégories classiques d'attribution causale.

L'interaction entre l'effet de la consigne et la perspective adoptée a également été testée sur les catégories classiques de la théorie d'attribution. Les catégories interne/externe suggère que l'asymétrie a un effet tendanciel, $F(1.772, 145.319) = 2.599, p < 0.1$. L'effet de la consigne n'est pas significatif, $F(1.772, 145.319) = 1.451 ; ns$. Toutefois, lorsqu'on observe les

moyennes des acteurs et des observateurs séparément, on remarque des variations différentes. Les acteurs, en situation d'autoprésentation utilisent de moins en moins d'explications externes, confirmant l'hypothèse d'un effet similaire à celui de la norme d'internalité, suggéré par Beauvois et Dubois (2009). 43% des explications d'acteurs renvoient à des explications externes dans la condition standard. Seul 32% des explications d'acteurs se réfèrent à des explications externes, dans la condition de désirabilité sociale. Le pourcentage descend à 24% en condition d'utilité sociale. Les observateurs, par contre, augmentent l'utilisation d'explications externes en condition de désirabilité. Le pourcentage débute à 23%, augmente à 30% dans la condition de désirabilité et redescend à 25% en condition d'utilité sociale. Ainsi, pour se faire apprécier, les observateurs, en moyenne, ont tendance à augmenter le nombre d'explication externe. Il se pourrait qu'un observateur imagine que pour se faire apprécier, il faille montrer de l'indulgence avec l'agent, en utilisant des explications externes. Nous ne pouvons toutefois pas écarter la possibilité que le contenu des réponses des observateurs n'ait pas été influencé par le fait qu'il est parfois difficile de s'autoprésenter lorsqu'on explique le comportement d'une tierce personne.

La figure 3 montre la variation des moyennes d'explication externe en fonction de la perspective et de la consigne.

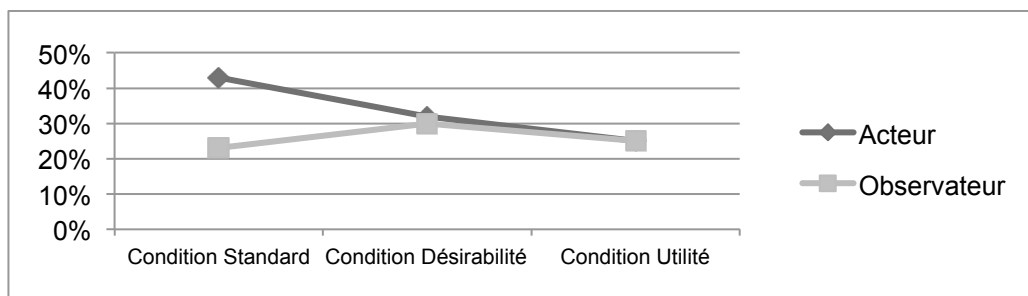


Figure 3. Moyennes des explications externes des acteurs et observateurs selon la consigne adoptée.

La catégorie des traits subit d'importantes variations en fonction de la condition et de la perspective. Les deux résultats étaient largement significatifs, $F(1.643, 136.343) = 6.794, p < 0.01$, $F(1.643, 136.343) = 7.770, p < 0.01$. Chez les acteurs, la moyenne d'utilisation de trait augmente en fonction de la consigne. (13% en condition standard, 20% en condition de désirabilité, 43% en condition d'utilité sociale). En condition d'utilité sociale, il semblerait que les acteurs augmentent considérablement leur utilisation en comparaison aux autres conditions, confirmant l'hypothèse de Cambon et al. (2006), suggérant que l'utilisation de trait est informative et permet de s'autoprésenter positivement dans les conditions d'utilité sociale. Chez les observateurs, la moyenne des traits varie également en fonction de la consigne. L'utilisation des traits baisse en condition de désirabilité sociale pour revenir se rétablir en condition d'utilité sociale.

En ce qui concerne l'asymétrie acteur-observateur, en condition standard, les différences entre les acteurs et les observateurs sont négligeables, mais l'écart grandit progressivement dans les conditions de désirabilité sociale et d'utilité sociale. Dans les conditions d'autoprésentation, les traits ne semblent pas aussi utiles pour les observateurs qu'ils ne le sont pour les acteurs.

La figure 4 présente un graphique des moyennes d'utilisation de traits selon la perspective et la consigne.

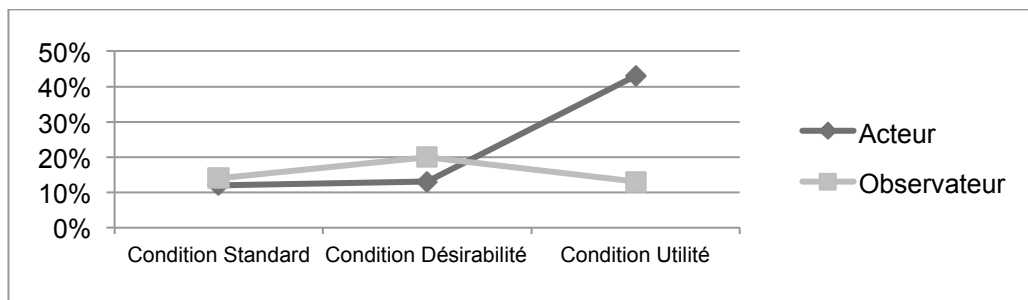


Figure 4. Moyennes d'utilisation de traits des acteurs et observateurs selon la consigne adoptée.

Discussion

Hypothèses du Modèle de Malle (2004)

L'objectif de ce travail était de tester les hypothèses de Malle et al (2007) avec les catégories provenant de la théorie naïve d'action intentionnelle et d'observer si l'introduction de variables, telles que la désirabilité et l'utilité sociale produirait des variations sur les explications naïves de comportement. Nous commencerons par discuter des quatre hypothèses de Malle dans la condition standard.

La première hypothèse se situe au niveau du choix du mode d'explication. Malle et al. (2000) prédit que les acteurs choisissent des comportements moins intentionnels que les observateurs. Dans la condition standard, nous avons effectivement remarqué que les acteurs citaient des comportements moins intentionnels que les observateurs. Malle et al. (2000) pensent que cette asymétrie s'explique par le fait que les comportements non-intentionnels soient plus intéressants à expliquer pour les acteurs que pour les observateurs.

La deuxième hypothèse avançait que les acteurs invoqueraient plus de raisons que de raisons historiques causales que les observateurs. Nos

résultats confirment partiellement cette prédiction. Les acteurs et les observateurs utilisaient le même pourcentage de raisons, mais les observateurs utilisaient plus de raisons historiques causales. Malle (2004) explique cette asymétrie par le fait que les observateurs auraient plus de mal à inférer les raisons directes qui ont causées le comportement. Cette explication semble plausible, mais comment expliquer la différence entre nos résultats et ceux de Malle ? Au moment du codage des réponses des participants, nous avons remarqué que les explications de notre échantillon différaient légèrement de celles utilisées en exemple dans les articles de Malle (2004, 2006). Le processus explicatif paraît être plus confus et dispersé chez nos participants qu'il ne l'était dans les exemples cités par Malle. Tout d'abord, les réponses de notre échantillon proposaient régulièrement un éventail de possibilités d'explication. Prenons un exemple de notre échantillon.

« Je n'avais plus sommeil. Mon lit me faisait mal au dos, j'étais contente de me réveiller pour une fois avant mon réveil et de pouvoir prendre le temps. J'avais envie de boire un thé, donc je me suis levée. »

Sur la base de cette explication, nous pouvons voir que les raisons qui ont poussé cette personne à se lever avant son heure de réveil sont nombreuses. Lorsque plusieurs modes d'explication sont mentionnés, Malle (2004) précise qu'il faut les coder individuellement, mais il n'indique pas quelle est la procédure à suivre lorsque l'on trouve plusieurs explications au sein d'un même mode d'explication. Nous pourrions à la fois considérer que dès lors que la personne mentionne des raisons, qu'il faille coder cette explication dans cette catégorie. Mais nous pourrions également décider de coder quatre

raisons. L'inconvénient est que cette dernière méthode pourrait fausser le pourcentage d'occurrences de raison à travers notre échantillon. Nous nous sommes donc limités à coder une fois en fonction des catégories citées. Nous avons également été confronté à une deuxième difficulté qui aurait pu affecter le résultat de cette hypothèse. A de nombreuses reprises, les participants commençaient déjà à apporter une explication au moment de la description de leur comportement. Ainsi, quand ils passaient à la question suivante, ils se sentaient peut-être obligés d'apporter une deuxième explication. Si l'on a déjà expliqué un comportement avec une raison, il paraît logique d'utiliser une raison historique causale comme deuxième explication. Il était difficile de savoir quelle explication coder. Fallait-il coder la première, la deuxième ou les deux ? Nous avons décidé de coder les deux. Ceci pourrait expliquer le grand nombre de raison historique causale. De plus, dans le cadre des études de Malle (1997b, 2000, 2007), les participants étaient des étudiants de son cours. Il se pourrait que ses participants aient été influencés par le contenu de son cours, comprenant qu'il attendait une seule explication.

La troisième hypothèse postulait que les observateurs préféraient les croyances aux désirs. Nos résultats indiquent que les acteurs et les observateurs choisissent plus de désirs que de raisons. D'après nous, les catégories de désir ou de croyance peuvent être difficile à différencier. Malle (2004) insiste sur une différence conceptuelle entre les deux catégories. Nous aurions envie d'argumenter que dans de nombreux cas, la différence se situe plus sur un niveau linguistique que conceptuel. Prenons un exemple.

« Je suis allé en Espagne parce que je voulais aller dans un pays chaud. »

« Je suis allé en Espagne parce que je pense qu'il fera chaud là-bas. »

Dans les deux cas, la personne exprime à la fois un désir et une croyance. Elle part en Espagne parce qu'elle pense qu'il fera chaud et qu'elle a envie d'aller dans un pays chaud. La différence entre les deux catégories est alors davantage linguistique. Si tel est le cas, alors ce problème conceptuel pourrait partiellement expliquer notre manque de résultats significatifs pour ces catégories.

La dernière hypothèse proposait que les observateurs utilisaient plus de marqueurs d'état mental que les acteurs. Nous n'étions pas en mesure de commenter cette prédiction en raison du nombre limité d'occurrences de marqueurs.

Sur la base de nos résultats, nous avons été en mesure de confirmer une seule des hypothèses de Malle (2004). La deuxième hypothèse est partiellement vérifiée, mais les deux dernières ne produisent pas les effets suggérés par Malle (2004). Les deux premières prédictions concernaient les modes d'explication (raison, cause, raison historique causale), alors que les deux dernières se rapportaient aux types d'explication (désir, croyance, et marqueur d'état mental). Nous envisageons la possibilité que ces résultats soient liés à la taille de notre échantillon. En effet, les modes d'explication étant des catégories plus larges, nous avons un nombre supérieur d'occurrence dans les catégories de mode d'explication que dans les types.

Au sein des catégories traditionnelles de la théorie classique d'attribution causale, seules les variables interne/externe ont produit un effet

d'asymétrie significatif dans la condition standard. Les acteurs et les observateurs ont tous deux produit plus d'explications internes, rejoignant les prédictions de Dubois et Beauvois (2008). Néanmoins, les acteurs ont utilisé plus d'explications externes que les observateurs dans la condition standard. L'asymétrie rejoint la théorie de Jones & Nisbett (1972) disant que les observateurs tendent à utiliser plus d'explications internes.

L'impact des Consignes d'Autoprésentation sur la Production Explicative

Nous allons nous atteler désormais à l'explication des résultats de la deuxième partie de notre étude portant sur l'effet de la consigne et de l'interaction de celle-ci avec l'effet d'asymétrie. Nous pouvons citer deux résultats dignes d'intérêt. Le premier concerne les variations de l'utilisation des raisons historiques causales. Les raisons historiques causales citées par les acteurs augmentent entre chaque condition. En condition d'utilité sociale, la moyenne d'occurrence de cette catégorie double chez les acteurs (voir figure 2). Si l'on s'en tient à l'explication de Dubois et Beauvois (2008), lorsque l'on place des personnes en situation d'évaluation, celles-ci peuvent être influencées par la volonté de s'autoprésenter positivement. Si l'on imagine que les enjeux d'évaluation sont plus importants pour les acteurs qu'ils ne le sont pour les observateurs, alors le postulat de Dubois et Beauvois (2008) est corroboré par le fait que l'utilisation des observateurs des raisons historiques causales ne subit pas d'importantes variations. Sur la base de nos résultats, on peut supposer que dans une situation évaluative, une personne utilise plus de raisons historiques causales. Pourtant, à suivre

Malle (2004), la fonction des raisons causales historiques est de combler un manque de connaissance. De plus, les explications de raison renvoient une image rationnelle de l'explicateur, d'après Malle (2004). Nous devrions donc trouver le résultat inverse. Quel pourrait être alors l'intérêt des raisons historiques causales ? Les raisons historiques causales sont composées d'explications impersonnelles, tirées de règles générales. Il est important de rappeler que Malle (2004) recommande de coder l'utilisation de traits de personnalité dans la catégorie des raisons historiques causales. Ces variations pourraient donc être simplement le produit d'une augmentation d'utilisation de trait dans les situations d'évaluation. Mais nous pourrions également envisager que les raisons historiques causales, tout comme les traits, apportent plus d'informations à l'évaluateur. En effet, invoquer la raison d'un comportement intentionnel pour un acteur n'est pas toujours informatif, il peut s'agir de raisons logiques et banales. A l'inverse, les raisons historiques causales peuvent informer l'évaluateur de la perception du monde d'un acteur. Le choix d'une règle générale plutôt qu'une autre (sur lesquelles sont basées les raisons historiques causales) informe l'évaluateur des opinions et valeurs de la personne. Prenons un exemple de notre échantillon. Dans la condition standard, la personne expliquait pourquoi celle-ci s'était fâchée avec sa mère. Elle explique : « C'était parce que je ressentais une aversion pour son comportement de préférence envers mon frère. ». Dans la condition de désirabilité sociale, elle explique son comportement de la manière suivante : « Il est difficile de grandir avec ses parents. Même si on est adulte, on reste marqué par des automatismes de fratrie. » La réponse est plus vague, mais elle reste personnelle. Dans la

condition d'utilité sociale, la personne répond de la manière suivante : « Les relations au sein d'une famille sont fortement liées à des patterns installés précocement. ». Cette réponse lui permet d'indiquer son point de vue sur les relations familiales et les « patterns » familiaux, sans devoir s'impliquer personnellement et de risquer un jugement personnel de l'évaluateur.

Le deuxième résultat qui nous a intéressé était celui indiquant une variation des catégories interne/externe et présence de trait/absence de trait (voir figure 3 et 4). Les acteurs utilisaient de plus en plus d'explications internes en situation d'évaluation. Cette variation n'était pas aussi importante chez les observateurs. Ces résultats pourraient confirmer l'existence d'une norme d'internalité, ce d'autant plus si l'on accepte l'hypothèse que les situations d'évaluations ont un impact plus important sur les acteurs que les observateurs. En effet, selon nos résultats, les variations n'étaient pas aussi importantes chez les observateurs. L'augmentation de traits peut aussi être expliquée par l'hypothèse qu'une des stratégies des évalués est d'apporter des explications informatives à l'évaluateur. Ceci expliquerait le fait qu'en condition standard, les acteurs citent des traits dans 13% des cas, alors qu'en condition d'utilité sociale, le pourcentage augmente à 43%.

Conclusion

Les prédictions de la théorie naïve des actions intentionnelles n'ont été que partiellement vérifiées. L'observation des résultats indique que certaines variables ont produit plus d'effet dans notre contexte. Les modes d'explication (les causes, les raisons et les raisons historiques causales) ont

apporté plus d'informations concluantes que les types d'explication (désir, croyance et marqueur d'états mentaux). La taille de notre échantillon a sans doute eu un effet sur ces prédictions. Les occurrences de désir ou de marqueur étaient vraisemblablement trop limitées pour faire apparaître des tendances significatives. De plus, les caractéristiques des types d'explication étant largement basées sur des considérations linguistiques, il se pourrait que la traduction ait été un enjeu important.

En dépit des suggestions de Malle (2006), les catégories traditionnelles de l'attribution causale ont produit les effets escomptés. Peut-on pour autant en conclure que la théorie des actions intentionnelles n'est pas pertinente pour interpréter les explications naïves de comportements ? La richesse de la théorie de Malle (2004) réside dans sa représentation complexe des différents niveaux d'analyse et des processus psychologiques qui dominent le développement d'explications naïves. Il est envisageable de penser que les catégories issues de la théorie naïve des actions intentionnelles et celles de l'attribution causale servent à des buts d'analyse différents. Les catégories interne/externe facilitent peut-être l'interprétation de la perception du poids causal de l'agent et de son impression de contrôle sur son environnement. La théorie naïve des actions intentionnelles rend compte d'autres phénomènes psychologiques qui déterminent la production d'explication. Les catégories interne/externe ne suffisent pas, par exemple, à déterminer l'influence des ressources informationnelles sur l'asymétrie acteur-observateur. De plus, le concept d'intentionnalité semble présenter théoriquement et empiriquement des suggestions dignes d'intérêt.

Quant aux hypothèses de Dubois et Beauvois (2008) les résultats semblent indiquer une présence de la norme d'internalité dans les explications analysées à l'aide catégories traditionnelles d'attribution causale. Comme prédit, les acteurs utilisaient plus de traits et d'explications internes dans les conditions évaluatives. La séparation du concept de valeur sociale en deux conditions semble également pertinente au vu de nos résultats. Nous pouvions distinguer des variations entre chacune des conditions, indiquant que l'explicateur adapte son explication selon sa volonté de se faire apprécier ou de paraître utile.

L'interaction des deux théories (celles de Malle et de Beauvois) a révélé des résultats intéressants. Certaines catégories, telles que les raisons, les raisons historiques causales et l'absence de marqueur ont été influencées par la condition. Nous ne disposons pas de preuves suffisantes pour affirmer que l'utilisation de la grille de lecture proposée par la théorie naïve d'action intentionnelle est un moyen entièrement adéquat d'interpréter les hypothèses de Beauvois. Néanmoins, l'effet de la condition sur l'utilisation des raisons historiques causales nous laisse penser que la manipulation des buts de gestion d'impression paraît être un moyen intéressant de faire apparaître des tendances au sein des catégories de la théorie naïve des explications intentionnelles. Dans ce contexte, l'utilisation des conditions de désirabilité et d'utilité sociale nous semble justifiée. Nous suggérerions que les conditions inspirées de Beauvois sont pertinentes dans le but de comprendre l'influence des stratégies d'autoprésentation dans la production d'explication naïve. Toutefois, nous ne pensons pas qu'elles soient bien adaptées à l'étude des asymétries acteur-observateur.

Références

- Baron, R. A., Byrne, D., & Branscombe, N. (2006). *Social psychology 11th ed.* Boston, MA Allyn & Bacon, MA.
- Beauvois, J., & Dubois, N. (1988). The norm of internality in the explanation of psychological events. *European Journal of Social Psychology, 18*(4), 299–316. doi: 10.1002/ejsp.2420180402
- Beauvois, J. L., & Dubois, N. (2009). Lay psychology and the social value of persons. *Social and Personality Psychology Compass, 3*(6), 1082–1095. doi: 10.1111/j.1751-9004.2009.00225.x
- Cambon, L., Djouari, A., & Beauvois, J. L. (2006). Social judgment norms and social utility: When it is more valuable to be useful than desirable. *Swiss Journal of Psychology/Schweizerische Zeitschrift für Psychologie/Revue Suisse de Psychologie, 65*(3), 167-180. doi: 10.1024/1421-0185.65.3.167
- Dompnier, B. (2006). *La valeur sociale des explications causales: norme d'internalité, jugements scolaires et registres de valeur* (thèse de doctorat non publiée). Université Pierre Mendès France, Grenoble, France.
- Dompnier, B., & Pansu, P. (2010). La Valeur Sociale des Explications Causales en Contexte Educatif. *Swiss Journal of Psychology, 69*(1), 39–51. doi: 10.1024/1421-0185/a000005
- Dubois, N. (1994). *La norme d'internalité et le libéralisme* (Vol. 9). Presses universitaires de Grenoble.
- Dubois, N., & Aubert, E. (2010). Valeur sociale des personnes: deux informations valent-elles mieux qu'une? *Revue internationale de*

psychologie sociale, (1), 57–92.

- Dubois, N., & Beauvois, J. L. (2008). The social value of internal explanations and the norm of internality. *Social and Personality Psychology Compass*, 2(4), 1737–1752. doi: 10.1111/j.1751-9004.2008.00126.x
- Fiske, S. T. (2009). *Social Beings: Core Motives in Social Psychology*. John Wiley & Sons.
- Gilibert, D., Cambon, L., & others. (2003). Paradigms of the sociocognitive approach. *A sociocognitive approach to social norms*, 5, 38-69.
- Gosling, P. (2007). Explication, norme d'internalité et norme de responsabilité. Dans J.-P. Pétard (dir.), *Psychologie sociale* (2nde édition revue) (pp. 269-334). Paris : Bréal.
- Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry*, 52(5), 449–466.
doi: 10.1080/00201740903302600
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Jellison, J. M., & Green, J. (1981). A self-presentation approach to the fundamental attribution error: The norm of internality. *Journal of personality and social psychology*, 40(4), 643-649.
doi: 10.1037/0022-3514.40.4.643
- Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79 –94).

- Morristown, NJ: General Learning Press.
- Jones, E. E., & Davis, K. E. (1966). From acts to dispositions The attribution process in person perception. *Advances in experimental social psychology*, 2, 219–266.
- Kassin, S., Fein, S., & Markus, H. (2008). *Social psychology 7th ed.* New York: Houghton Mifflin.
- Knobe, J., & Malle, B. F. (2002). Self and other in the explanation of behavior: 30 years later. *Psychologica Belgica*, 42(1/2), 113–130.
- Malle, B. F. (2004). *How the mind explains behavior: Folk explanations, meaning, and social interaction.* The MIT Press.
- Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, 132(6), 895.
doi: 10.1037/0033-2909.132.6.895
- Malle, B. F. (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. *Advances in experimental social psychology*, 44, 297–352.
- Malle, B. F., & Knobe, J. (1997a). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33, 101–121.
- Malle, B. F., & Knobe, J. (1997b). Which behaviors do people explain? A basic actor–observer asymmetry. *Journal of Personality and Social Psychology*, 72(2), 288–304. doi: 10.1037/0022-3514.72.2.288
- Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, 93(4), 491–514. doi: 10.1037/0022-3514.93.4.491

- Malle, B. F., Knobe, J., O’Laughlin, M. J., Pearce, G. E., & Nelson, S. E. (2000). Conceptual structure and social functions of behavior explanations: Beyond person–situation attributions. *Journal of Personality and Social Psychology*, *79*(3), 309-326.
doi: 10.1037/0022-3514.79.3.309
- Malle, B. F., & Nelson, S. E. (2003). Judging mens rea: The tension between folk concepts and legal concepts of intentionality. *Behavioral sciences & the law*, *21*(5), 563–580. doi: 10.1002/bsl.554
- Monroe, A. E., & Malle, B. F. (2010). From uncaused will to conscious choice: The need to study, not speculate about people’s folk concept of free will. *Review of Philosophy and Psychology*, *1*(2), 211–224.
doi: 10.1007/s13164-009-0010-7
- Shaver, K. G. (1975). *An introduction to attribution processes*. Cambridge, MA: Winthrop
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in experimental social psychology*, *21*, 261–302.
- Watson, D. (1982). The actor and the observer: How are their perceptions of causality divergent? *Psychological Bulletin*, *92*(3), 682-700.
doi: 10.1037/0033-2909.92.3.682