L’approbation sociale et l'estime de soi sont des éléments fondamentaux qui guident les actions humaines. La connaissance de ce qui est socialement acceptable et de ce qui ne l’est pas se développe très tôt chez les enfants (Baron-Cohen 1991; Call et Tomasello 1998 ; Meltzoff 2002 ; Decety et Sommerville 2003). Dans toutes les cultures, à n’importe quel moment de leur histoire, les individus ont toujours pondéré leurs actions en fonction de la façon dont celles-ci seront perçues par les autres, sacrifiant parfois beaucoup pour maintenir réputation, honneur et respect. Non seulement les individus se soucient de l'opinion d’autrui, mais ils se préoccupent également de leur propre opinion sur eux-mêmes ; se comporter selon la morale, se sacrifier pour un idéal ou un principe, agir «comme si» nous étions toujours sous le jugement d’autrui, sont autant d’exemples de mécanismes mis en place par les individus pour maintenir l’image et l’estime de soi.

Les décisions économiques ne font pas exception. Adam Smith a été l'un des premiers, et pour longtemps le seul économiste, à reconnaître l'importance de l'image dans les échanges économiques. Dans sa Théorie des Sentiments Moraux (1759), Smith met en évidence le désir des gens de préserver et même améliorer leur image sociale, et il montre comment les individus ont tendance à comprendre leurs actions comme le moyen d’évaluer et soumettre à l’épreuve leurs vertus, pour eux-mêmes et pour leurs pairs. Selon lui :

« L'homme ne désire pas seulement être aimé : il désire aussi être aimable, ou être digne de ce qui mérite véritablement l’amour. Il craint à la fois d’être haï et d’être haïssable, ou de justifier véritablement la haine. Il désire non seulement la louange, mais aussi la louange méritée, c’est-à-dire, être un objet naturel de louange quand même il ne serait loué par personne. Il craint non seulement le blâme, mais aussi le blâme mérité, c’est-à-dire, être un objet naturel de blâme quand bien même il ne serait blâmé par personne »(The Theory of Moral Sentiments, partie III, Ch. 2, p.xx).

L’idée essentielle de Smith était que, tandis que les gens désirent la louange et la reconnaissance, ils comprennent néanmoins que leurs propres actions doivent être conformes afin de dégager une *utilité* de la louange. Pour paraphraser Smith, les individus sont guidés dans leurs actions par la métaphore d’un spectateur impartial qui, dans un processus récursif et endogène, influence à la fois les jugements des actions des autres et les jugements et les actions des individus eux-mêmes.

Bien que les philosophes, les historiens, les psychologues et les premiers économistes comme Smith aient montré de l’intérêt pour le rôle de l'image de soi dans les décisions humaines, les économistes l’ont ignoré pendant longtemps. La raison de cette désaffection est à la fois philosophique et méthodologique.

D'un point de vue philosophique, le travail des premiers économistes comme Adam Smith a été en quelque sorte mal compris : tandis que la Richesse des Nations devenait le texte de référence de la pensée de Smith, sa notion complexe et multiforme d'*intérêt* a été interprétée selon le concept plus pragmatique d’intérêt monétaire : *l'homo oeconomicus* était né. C'est seulement au cours du siècle dernier

que les économistes ont commencé à (re)développer la notion d'intérêt pour étudier des comportements qui seraient difficiles à expliquer par le paradigme de *l'homo* oe*conomicus* : les dons de bienfaisance, la fourniture de biens public, le bénévolat, l'honnêteté fiscale, la discrimination, l'acceptation de l'autorité des supérieurs au sein des entreprises, les salaires d'efficience, sont autant d'exemples de décisions stratégiques que l'on expliquerait difficilement en limitant la notion d'intérêt à un intérêt purement monétaire. Les contributions d'économistes comme Gary Becker et Kenneth Arrow, et le développement de domaines pluridisciplinaires tels que l'économie comportementale et expérimentale témoignent de l'importance des facteurs non monétaires dans la prise de décision économique.

D'un point de vue méthodologique, l'introduction de la cognition sociale, de l'image de soi et des croyances dans le discours économique a été rendue possible par des innovations importantes au niveau des outils théoriques et mathématiques, ainsi que des méthodes empiriques. Sur le plan théorique et mathématique, le développement de la théorie des jeux et des statistiques bayésiennes ont augmenté notre capacité de générer des prédictions théoriques sur la formation des croyances et sur les interactions stratégiques. Sur le plan empirique, la révolution dans la statistique moderne commencée par Ronald Fisher et l'introduction de l'expérimentation dans l'économie ont rendu possible de générer et vérifier des hypothèses qu'il serait difficile de tester avec des données naturelles. En outre, depuis quinze ans, la neuroéconomie, un nouveau champ interdisciplinaire qui combine les expériences économiques avec des techniques de neurosciences, a élargi notre connaissance de la façon dont les individus évaluent et traitent les récompenses monétaires et non monétaires. Cette exploration est particulièrement importante pour analyser les décisions économiques impliquant une motivation en termes d'image. En effet, la compréhension de la façon dont le cerveau calcule les gains et pertes d'images dans différentes situations aide à discerner pourquoi, tandis que certaines politiques et pratiques visant à encourager des comportements pro-sociaux sont efficaces, d'autres ne le sont pas.

Comprendre comment les différents problèmes de décision sont affectés quand la fonction d'utilité incorpore un facteur d'image est important non seulement d'un point de vue philosophique, mais aussi d'un point de vue empirique. Puisque agir pro-socialement est considéré comme une caractéristique souhaitable, les politiques publiques et les pratiques privées génèrent des incitations ayant comme objectif l'encouragement à de telles décisions : les bénévoles sont encouragés à donner avec des mécanismes de contrepartie, les gouvernements incitent les dons de bienfaisance en permettant des déductions fiscales, ils subventionnent les produits respectueux de l'environnement (par exemple, véhicules à faibles émissions), et ils utilisent des mécanismes de répression pour décourager des comportements antisociaux comme l'évasion fiscale ou les comportements anti-environnementaux. Même si ces pratiques sont destinées à créer des incitations personnelles (souvent monétaires) qui compensent les coûts supportés par les individus pro-sociaux, elles peuvent parfois devenir contreproductives. De même, des conséquences inattendues peuvent se produire alors que les gouvernements ou des entités privées

essayent d'utiliser la pression sociale comme outil d'incitation aux comportements pro-sociaux. Benabou et Tirole (2006), par exemple, ont fait l'hypothèse que, bien que certaines personnes soient sincèrement altruistes, d'autres voient de bonnes actions telles que les dons de bienfaisance comme un investissement en termes d'image sociale. Ces investissements sont réalisés pour établir ou maintenir l'estime sociale et maintenir sa propre idée de soi. Pour ces individus, la présence d'incitations extrinsèques (monétaires) de même qu'une plus grande visibilité des actions changent le sens attaché aux comportements pro sociaux. Comme ce mécanisme incitatif attire des personnes plus matérialistes, la valeur de signal des actions pro-sociales se dilue.

Tout aussi important est de comprendre comment les décisions stratégiques sont influencées par des préoccupations d'image ou de fierté. Au cours des vingt dernières années, plusieurs études empiriques ont développé l'idée que les travailleurs ne sont pas uniquement motivés par l'intérêt pécuniaire, mais considèrent leur travail comme une source de réalisation de soi. En conséquence, fournir des incitations monétaires ou contrôler ses subordonnés peut occasionner des coûts cachés. Il a été démontré que l'augmentation des salaires induit parfois de meilleures performances (Fehr, Kirchsteiger et Riedl 1993; Gneezy et Rustichini 2000) mais que l'introduction de plus fortes incitations monétaires parfois induisent au contraire de plus mauvaises performances (Frey et Oberholzer-Gee, 1997; Fehr et Rockenbach 2003; Fehr et List 2004). De même, l'utilisation du pouvoir au sein des organisations a des effets comportementaux importants qui seraient difficiles à interpréter sans prendre explicitement en compte les préoccupations en termes d'image de soi ou d'orgueil (Falk et Kosfeld 2006). Les supérieurs ont la tendance à conserver la totalité de leur pouvoir de décision même si ce choix est financièrement sous-optimal par rapport à la délégation (Fehr, Herz et Wilkening 2013). Les subordonnés s'opposent souvent au contrôle pour des raisons autres que les incitations financières (par exemple la distance sociale, le dépit de n'être pas reconnu digne de confiance, l'autorité illégitime, etc.). Beaucoup de ces interrogations trouvent une réponse si l'on prend explicitement en compte le fait que les gens se soucient de leur autoréalisation (Benabou et Tirole 2006), de leur estime de soi et de leur orgueil (Ellingsen et Johannesson 2008).

Les trois essais présentés ici s'intéressent à deux domaines de prise de décision qui ont un lien important avec l'image sociale et l'image de soi : les dons charitables et les rapports hiérarchiques dans la firme. Le premier essai vise à déterminer si la pression sociale fonctionne de la même manière dans des décisions qui impliquent faire le bien ou éviter de faire le mal, en étudiant à la fois les réponses comportementales et neuronales à la pression sociale. Le deuxième essai répond à la question de savoir si les donateurs tiennent à la qualité des bénéficiaires et explore la relation entre la qualité et la quantité du don pour les donateurs motivés par le prestige. Le troisième essai étudie les effets de l'autorité dans les relations hiérarchiques caractérisées par des intérêts monétaires alignés entre un principal et un agent, en explorant comment différentes façons d'exercer le contrôle

interagissent avec l'image de soi à la fois des subordonnés et des supérieurs hiérarchiques.

Les raisons pour lesquelles les individus sacrifient des ressources personnelles en faveur du bien-être d'autrui ont été largement étudiées dans la littérature économique. Puisque l'idée traditionnelle que les gens sont motivés seulement par un pur altruisme n'est pas à même d'expliquer plusieurs observations empiriques sur les dons de bienfaisance, différentes théories alternatives ont été proposées pour expliquer la générosité privée. Suivant l'idée que les individus peuvent tirer une utilité directe de l'acte de donner (voir par exemple Becker, 1974), des facteurs comme les normes intériorisées (Arrow, 1971; North, 1981), l'approbation sociale (Hollander, 1990), le warm-glow (Andreoni 1990 ; Ribar et Wilhelm, 2002; Harbaugh , Mayr et Burghart 2007), la coopération conditionnelle (Fischbacher, Gächter, et Fehr, 2001), la réciprocité (Sugden, 1984), et le prestige (Harbaugh, 1998a, 1998b; Bracha, Heffetz et Vesterlund 2009), ont été identifiés comme des moteurs puissants du don. En particulier, les données empiriques sur l'importance de l'image sociale et le prestige sont maintenant nombreuses (Vesterlund 2003 ; Andreoni et Petrie, 2004 ; Rege et Telle 2004; Ariely, Bracha et Meier 2009 ; Shang et Croson 2009).

Les deux premiers essais contribuent à cette littérature.

Dans le premier essai, nous utilisons l'imagerie par résonance magnétique (IRMf) pour étudier les corrélations neuronales dans deux types de décisions pro-sociales: *faire le bien* et *refuser de faire le mal*. En particulier, nous étudions deux types de situations où les récompenses morales et monétaires sont de sens opposé : (1) les gens peuvent subir une perte monétaire pour permettre un don à une organisation bien évalué (don coûteux pour une « bonne » association), et (2) peuvent refuser de gagner de l'argent afin d'éviter un transfert monétaire vers une organisation évaluée négativement (opposition coûteuse à un transfert).

Les informations sur l'activité neurale pendant la prise de décision sont particulièrement importantes. Elles permettent de comprendre dans quelle mesure et pourquoi les gens sont prêts à troquer des pertes monétaires contre des gains d'image sociale.

En faisant varier systématiquement le coût du don et le degré de visibilité sociale de l'action, nous montrons que pour ces deux types de décisions, les récompenses morales et monétaires sont pondérées et évaluées par différentes régions du cerveau, selon que l'image sociale est une préoccupation ou non. Pour faire cela, nous avons implémenté un protocole 2x2 administré aux mêmes sujets où les individus décident d'accepter ou de rejeter des transferts à deux organismes différents, en présence ou pas d'observateurs. Pour les choix concernant l'association positivement évaluée, les sujets décident d'accepter ou pas des dons faits par l'expérimentateur à l'organisation, à un coût variable pour eux-mêmes prélevé sur leur dotation initiale. Dans les choix concernant l'association

négativement évaluée au contraire, les sujets décident d'accepter ou pas des gains différents pour eux-mêmes, qui sont associés en contrepartie à divers transferts monétaires effectués par l'expérimentateur à l'association négativement évaluée. Dans ce protocole par conséquent, la seule façon de gagner de l'argent pour un sujet est de laisser l'expérimentateur transférer de l'argent à une organisation évaluée négativement, alors que tout transfert à un organisme évalué positivement entraîne une perte monétaire. Cette configuration nous permet de déterminer si, pour ces deux types de décisions, les récompenses morales et monétaires sont pondérées et évaluées dans le cerveau de la même manière, et comment les préoccupations sociales en termes d'image influent sur le comportement et ses corrélats neuronaux. Nos résultats montrent que *faire le bien* et *éviter de faire le mal* impliquent des processus neuronaux partiellement différents et que l'effet de l'exposition sociale sur la décision dépend à la fois de la nature de la décision pro-sociale et des gains ou coûts en cause.

Nous constatons que les gains en termes d'image sociale résultant d'une décision publique généreuse et les gains monétaires dérivant d'un choix égoïste « en privé » ne partagent pas les mêmes réseaux de neurones et diffèrent partiellement : les gains d'image sociale sollicitent pour les deux décisions le Cortex Cingulaire Antérieur et le putamen (régions connues pour traiter les conflits et évaluer les événements liés aux récompenses) ; les gains monétaires dérivant de décisions égoïstes prises en privé sont dissociés dans le Cortex Orbitofrontal (OFC) : refuser de faire un don est en corrélation avec l'OFC médian droit, une région connectée à l'appréciation de la valeur des récompenses (par exemple l'argent) ; accepter de gagner de l'argent et permettre un transfert à une association négativement évaluée corrèle avec l'OFC latérale, une région impliquée dans l'évaluation des punitions. De plus, nous montrons que l'exposition sociale a peu d'impact quand faire le mal comporte des pertes morales faibles (par exemple gains faibles pour l'association négativement évaluée), tandis que pendant ces essais l'activité dans l'IPC et le Cortex Frontopolaire est plus élevée lorsque les décisions sont prises à huis clos. Quand il s'agit de faire le bien, au contraire, le cortex pariétal inférieur se trouve être plus activé en privé qu'en public chaque fois que les gains potentiels pour l'association sont élevés. Le rôle connu de ces régions dans l'élaboration du sens individuel de l'agence soutient l'interprétation que lorsque des décisions égoïstes ont des conséquences négatives limitées pour autrui, l'exposition sociale est inefficace pour corriger des comportements qui sont subjectivement perçus comme négatifs mais finalement "sans danger" ou sans véritable conséquence sérieuse.

Pris ensemble, ces résultats ont des implications importantes pour les politiques publiques : lorsque l'objet est d'inciter les gens à faire du bien, notre travail confirme que les politiques visant à accroître la pression sociale sur les individus sont efficaces. Toutefois, lorsque l'objectif d'une politique est de dissuader des comportements qui génèrent des externalités sociales négatives, notre expérience suggère que l'efficacité de la pression sociale dépendra de l'ampleur des dommages créés : dans notre expérience, la visibilité sociale ne dissuade ces comportements négatifs que quand les décisions impliquent un transfert élevé à l'association

négativement évaluée. Pour toute autre combinaison de gains personnels et de gains pour l'association négativement évaluée, au contraire, l'effet de l'exposition sociale est négligeable, même lorsque les gains anticipés pour lui par l'individu sont faibles. Nos résulats donnent un aperçu de la façon dont l'exposition sociale affecte différents types de comportements pro-sociaux. Alors que la pression sociale est efficace lorsqu'il s'agit de faire le bien, son impact sur les décisions qui évitent de faire le mal est limité. Cela peut aider à expliquer pourquoi il est difficile de changer les habitudes des individus quand il s'agit de choix qui, bien que nuisant à la société au niveau agrégé, sont difficilement perçus à un niveau individuel comme «répugnant» (par exemple ne pas recycler, traverser la rue au feu rouge, ne pas se faire vacciner contre la grippe, polluer).

Alors que le premier essai explore la relation entre les problèmes d'image et de la quantité du don, le deuxième essai étudie comment la générosité des donateurs est affectée par des informations sur la qualité des bénéficiaires du don et comment cette relation est affectée par la prise en compte de préoccupations d'image.

Dans la dernière décennie, la quantité d'informations disponibles sur la responsabilité financière et l'efficacité des organismes de bienfaisance a augmenté de façon spectaculaire. Beaucoup de « veilleurs d'associations » sont en train de développer des mesures synthétiques visant à aider les donateurs à comparer des organismes de bienfaisance hétérogènes et à prendre des décisions plus éclairées. Charity Navigator, GiveWell, et Urban Institute sont des exemples de ces veilleurs à but non lucratif. La disponibilité de ces plate-formes a des conséquences importantes sur le marché des dons charitables, surtout quand il s'agit de petits donateurs. Alors qu'en fait les gros donateurs ont toujours eu à la fois les incitations et les moyens de recueillir ce type d'information, déterminer la solvabilité d'une association donnée exigeait jusqu'à aujourd'hui des efforts de recherche non triviaux pour les petits donateurs : collecter des documents gouvernementaux pertinents, appeler l'association pour demander des renseignements, interviewer les administrateurs et membres du conseil, ou même siéger directement dans les conseils. L'émergence de ces plate-formes d'information incite maintenant les organismes de bienfaisance à rendre compte de leur propre initiative, puisque ne pas être évalué pourrait sembler suspicieux aux donateurs. Comme ces mécanismes augmentent la quantité d'information disponible et réduisent le coût d'accès à l'information, les donateurs de tous niveaux peuvent désormais s'informer à relativement peu de frais. Alors que certains donateurs peuvent bien entendu choisir de rester dans l'ignorance, l'effet que cette information a sur ceux qui la prennent en compte est encore incompris.

L'objectif de ce deuxième essai est donc double: (i) étudier comment les informations sur l'efficacité réelle des organismes de bienfaisance affectent les montants des dons des petits donateurs ; et (ii) comprendre si la visibilité publique de cette information est importante pour les donateurs, c'est-à-dire si les donateurs motivés par l'image sociale prennent en compte l'information, sachant que même si elle peut être ignorée, elle ne peut pas être cachée aux autres. Autrement dit, nous

examinons si pour les petits donateurs, la *taille du don* et l'*efficacité* de l'association sont des compléments ou des substituts, et comment la réponse à l'information dépend du fait que *l'information* est publique ou non.

Dans ce but, nous avons mis en œuvre une expérience en deux phases, la seconde phase étant inconnue des sujets jusqu'à la fin de la première phase. Lors de la première phase, les sujets choisissent librement trois associations charitables parmi une liste de plus de 5000 évaluées par Charity Navigator. Ils jouent ensuite trois jeux du dictateur indépendants avec les organismes sélectionnés. Les sujets sont également autorisés à indiquer une association préférée parmi les trois, ce qui augmente la probabilité que cette association soit tirée au sort pour le paiement final. Alors que la première phase nous sert à évaluer la volonté de donner inconditionnelle des sujets, dans la deuxième phase nous évaluons comment les décisions prises dans la première phase sont révisées à la lumière des nouvelles informations sur les organismes de bienfaisance choisis. Dans la deuxième phase, les sujets sont incités à révéler ce qu'ils croient être la véritable efficacité de leurs associations, et ils reçoivent ensuite en privé les informations sur l'efficacité réelle de ces associations. Nous considérons que découvrir qu'une association est plus efficace qu'anticipé constitue une « bonne nouvelle », alors que découvrir qu'une association est moins efficace qu'anticipé est considéré comme une « mauvaise nouvelle ». Enfin, les sujets ont la possibilité de réviser leurs décisions initiales de dons et d'indiquer une nouvelle association préférée s'ils le souhaitent. Cela signifie que les participants peuvent, d'une phase à l'autre, soit changer leur association préférée, soit indiquer une favorite dans la phase 2 alors qu'ils n'avaient rien indiqué en phase 1 (et vice versa). Il est important de noter que les décisions de la phase 2, qui sont les seules prises en compte pour les paiements, ne peuvent pas influencer celles qui sont prises dans la phase 1 par définition, car les sujets au cours de la phase 1 ignorent l'existence d'une phase ultérieure. Cela signifie également que les participants ne savent pas si les autres ont changé ou pas leurs décisions entre les phases. Nous comparons les réponses à l'information sur l'efficacité des associations avec différents niveaux d'exposition sociale, et nous mettons en œuvre trois traitements : (T0) les décisions de dons et l'efficacité sont des informations privées, (T1), le montant du don effectué à l'association en 2^{e} phase est révélé publiquement mais l'information sur l'efficacité de l'association reste privée, et (T2) à la fois l'efficacité de l'association et le montant du don effectué à l'association en 2^{e} phase sont révélés publiquement.

Nos résultats montrent que le type d'information et son degré de visibilité ont des conséquences importantes sur le comportement des donateurs. Nous constatons que dans la mesure où l'efficacité des associations reste une information privée, les individus récompensent les associations plus efficaces que prévu (bonnes nouvelles) en augmentant leurs dons. Quand les associations de bienfaisance sont moins efficaces que prévu, les individus ne réduisent pas leurs dons quand toute l'information est privée. Toutefois, lorsque l'efficacité de la charité est révélée à des tiers, certaines associations reçoivent moins qu'en première phase alors que les nouvelles étaient bonnes, et certaines reçoivent plus qu'en première phase bien que

les nouvelles aient été mauvaises. Ce nouveau comportement, virtuellement absent dans les deux autres traitements, concerne 30% des sujets qui ont fait un changement de don en réponse aux nouvelles informations reçues. Nous suggérons que ce comportement est imputable aux donateurs motivés par leur image sociale. Ces derniers traiteraient la quantité du don et la qualité du bénéficiaire comme *substituts* en termes de gain d'image sociale. Ils réaffecteraient donc leurs dons afin de maintenir le même gain d'image espéré avant la deuxième phase de l'expérience.

Pris ensemble, ces résultats montrent deux effets importants de l'information sur le marché des dons : dans la mesure où l'information sur l'efficacité des organismes de bienfaisance reste privée (dans le sens où un donneur motivé par l'image sociale peut annoncer ses dons sans signaler implicitement l'efficacité du bénéficiaire), la diffusion de mesures synthétiques de qualité des associations augmenterait la taille du marché pour les organismes de charité les plus performantes en termes d'efficacité. Toutefois, dès que l'information est livrée avec les noms des organismes, paradoxalement ceux qui ont des résultats supérieurs en termes d'efficacité risquent de perdre une partie de leur financement à cause de donateurs motivés par leur image sociale qui sont prêts à échanger le montant de leur don avec la qualité du bénéficiaire. Autrement dit, avec l'information sur l'efficacité rendue disponible pour tous, les donateurs motivés par le prestige traiteraient la quantité de leurs dons et la qualité des bénéficiaires comme des *substituts*.

Alors que les deux premiers essais se concentraient sur des décisions pro-sociales et leur lien avec la valeur que les individus attribuent à leur image sociale, le troisième essai explore le rôle de l'image de soi dans les relations hiérarchiques au sein d'une organisation. Nous étudions dans un environnement d'entreprise comment différentes institutions hiérarchiques affectent l'acceptabilité de l'autorité formelle des supérieurs hiérarchiques par les subordonnés.. Plus précisément, nous examinons si l'utilisation de la part des agents et leurs supérieurs de leurs autorités formelles et réelles (au sens d' Aghion et Tirole 1997) dépend de la valeur de signal des restrictions de liberté de choix et de la délégation. L'objectif de l'essai est d'étudier des environnements où les supérieurs et les agents font face à des intérêts monétaires alignés. Il s'agit de comparer les situations où l'autorité des supérieurs est exercée impersonnellement (par exemple les règles s'appliquent à tous les subordonnés, quel que soit leur comportement), avec des situations où l'autorité est exercée par surveillance directe (par exemple, le supérieur impose des restrictions ad hoc lorsqu'il le juge nécessaire). L'enjeu est donc de comprendre si l'exercice du contrôle sur les subordonnés comporte des coûts cachés même si les préférences monétaires des supérieurs sont parfaitement alignées avec celles des subordonnés. Nous faisons l'hypothèse que si les agents ont des préférences sur le type de signal qu'une restriction comporte, alors ils devraient s'opposer à l'autorité de leurs supérieurs quand celle-ci est exercée avec un contrôle direct. En effet, cela signale que le supérieur est en désaccord avec leurs choix et qu'il est cherche à limiter l'autorité réelle des subordonnés.

La théorie de l'agence traditionnelle prédit que lorsque les intérêts monétaires des agents et des supérieurs ne sont pas alignés, les agents agissent dans leur propre intérêt, faisant de l'exercice du contrôle un moyen pour les supérieurs de limiter les décisions des agents opportunistes. De récentes analyses empiriques ont cependant remarqué que les supérieurs pourraient faire face à des coûts cachés, dès qu'ils essaient de contrôler leurs agents. Les agents qui sentent la méfiance à leur égard (Falk et Kosfeld 2006) ou perçoivent l'autorité du principal comme illégitime (Schneidler 2010) semblent en fait sanctionner le contrôle en fournissant moins d'efforts que si les supérieurs avaient délégué. Plus généralement, la distance sociale (Frey 1993; Dickinson et Villeval 2008), les intentions qui motivent les décisions des supérieurs (Charness et Levine 2007) et les menaces de punition (Fehr et List 2003) constituent autant de facteurs importants qui déterminent le résultat final, soit positif, soit négatif, du contrôle.

Dans cet essai, nous testons l'hypothèse que le contrôle peut recéler des coûts cachés, même lorsque les intérêts monétaires du mandant et du mandataire sont alignés. Nous introduisons un nouveau jeu d'investissement impliquant une triade supérieur-agent-client. Le contrôle prend la forme d'une limitation de la réciprocité possible de l'agent à l'égard d'un tiers (par exemple un investisseur ou un client) qui fait confiance à l'agent. Le supérieur peut en effet exercer une autorité formelle sur l'agent (par exemple en fournissant une limite supérieure à ce que l'agent peut renvoyer au client), tandis que l'agent conserve le pouvoir réel (par exemple, il met en œuvre la décision dans les limites autorisées). Puisque le montant transféré par le client et augmenté par l'expérimentateur qui n'est pas envoyé au client est réparti de manière égale entre l'agent et le supérieur, ces deux derniers ont des intérêts monétaires alignés. Nous faisons varier la capacité de surveillance des supérieurs vis-à-vis de l'agent. Nous évaluons ainsi comment la réciprocité des agents vers leurs clients dépend du fait que l'autorité formelle du supérieur représente une restriction personnelle ou une règle impersonnelle (appliquée ex ante à tous les types d'agents) qui signale simplement les préférences du supérieur.

Nos résultats montrent que même avec des intérêts monétaires alignés, le contrôle direct implique des coûts cachés : les agents restreints sont relativement plus généreux avec leur client (et donc moins généreux avec leur supérieur et avec eux-même) lorsque la restriction est personnelle (le principal fait du contrôle direct) plutôt qu'impersonnelle (le principal n'est pas informé des actions spécifiques du subordonné). Ces coûts cachés du contrôle peuvent être atténués en exerçant le contrôle à travers des règles impersonnelles. Plus précisément, les agents semblent être prêts à perdre de l'argent pour punir le principal en envoyant plus d'argent au client chaque fois que la restriction imposée par celui-ci représente une restriction intentionnelle (par exemple, « je te contrains parce que *tu as fait* mal ») ; par contre quand une restriction signale seulement les préférences du supérieur (par exemple, « je te contrains parce que *je préfère gagner davantage* »), les agents réduisent de manière significative leur transferts vers leur client, augmentant ainsi les gains à la fois du supérieur et d'eux-même.

Étonnamment toutefois, bien que les gains soient relativement plus élevés avec des règles impersonnelles, nous constatons que les supérieurs ont tendance à recourir à leur autorité formelle relativement moins souvent avec un système de règles impersonnelles. Nous supposons que les supérieurs sont réticents à révéler en premier leurs préférences égoïstes,et préférent ne signaler leur avarice que lorsque cela est strictement nécessaire : si, en fait, les supérieurs ont des préférences pour paraître justes et dignes de confiance (mais pas nécessairement d'être juste et digne de confiance) (voir exemple Dana, Weber, Kuang 2007 ; Hao et Houser 2011), alors des restrictions personnelles leur permettent de signaler leur appât du gain seulement quand cela est nécessaire (par exemple, quand l'agent est trop généreux avec le client). Au contraire, imposer à tous une règle impersonnelle signale quel type le supérieur est, aussi à des agents qui n'avaient pas besoin de savoir, car ils étaient déjà en train de prendre la « bonne » décision. Ce résultat semble être en accord avec des études récentes en matière de délégation et d'aversion à la culpabilité. Bartling et Fischbacher (2012) et Grossman (2012) ont montré comment la délégation peut être un instrument efficace que les supérieurs peuvent utiliser pour déplacer la culpabilité et la responsabilité de la punition sur leurs employés. Cet essai complète cette littérature en montrant comment la délégation de choix inégalitaires peut être important pour les supérieurs, même quand ils ne font pas face au risque d'être puni par les clients, et quand leur responsabilité ne peut pas être manipulée.