

Comportements (non) éthiques et stratégies morales

Marie Claire Villeval

▶ **To cite this version:**

Marie Claire Villeval. Comportements (non) éthiques et stratégies morales. Revue Economique, 2019, 70 (6), pp. 1021-1046. 10.3917/reco.706.1021 . halshs-02445185

HAL Id: halshs-02445185
https://shs.hal.science/halshs-02445185

Submitted on 19 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Published in *Revue Economique*, special issue Prix Revue Economique 2018, 70 (6), 1021-1046, 2019.

Comportements (non-)éthiques et stratégies morales

Marie Claire Villeval *

La théorie du crime de Becker a été mise à mal par une accumulation de preuves expérimentales qui ont mis en évidence une aversion intrinsèque des agents économiques au mensonge et l'importance de leurs croyances sur la perception de leur conduite morale par autrui. Cherchant à préserver leur réputation et leur image quand leurs actes génèrent des externalités négatives, les individus développent une variété de stratégies leur permettant de motiver leurs croyances et accroître l'incertitude, générant des limitations cognitives motivées. Nous illustrons ces stratégies à travers un ensemble de recherches récentes qui témoignent de l'émergence d'une véritable théorie comportementale des choix moraux alternative.

UNETHICAL BEHAVIOR AND MORAL STRATEGIES

Becker's theory of crime has been undermined by an accumulation of experimental evidence that has highlighted an intrinsic lying aversion of economic agents and the importance of their beliefs on the perception of their moral conduct by others. Seeking to preserve their reputation and image when their actions generate negative externalities, individuals develop a variety of strategies to motivate their beliefs and increase uncertainty, generating motivated cognitive limitations. We illustrate these strategies through a set of recent research that testifies to the emergence of an alternative behavioral theory of moral choices.

Classification *JEL* : C9, D91, K42.

* CNRS. Correspondance : Groupe d'Analyse et de Théorie Économique (GATE), 93 chemin des Mouilles 69130 Écully, France. Courriel : villeval@gate.cnrs.fr

Cet article a bénéficié des commentaires reçus lors de la journée de la Revue Économique du 11 janvier 2019. Je suis également reconnaissante à mes co-auteurs pour les nombreuses discussions qui ont permis d'enrichir ce texte.

INTRODUCTION

La fraude et la tricherie sont massivement répandues. Au-delà des exemples anecdotiques donnés très régulièrement dans les médias, on estime que la fraude fiscale se monte à environ 18 % des revenus à déclarer aux Etats-Unis. Dans le monde, la corruption et les pots de vins représenteraient environ 3% du PIB mondial. La fraude touche tous les domaines. Dans le secteur financier aux Etats-Unis, 7% des conseillers ont été sanctionnés pour mauvaise conduite, et le taux monte même à 15% dans certaines banques (Egan *et al*. [2019]). Dans l'éducation, il a été montré que l'introduction de tests à enjeux élevés a poussé les enseignants à frauder (*e.g*., Jacob and Levitt [2003] ; Martinelli *et al*. [2018]) et plus de 65% des étudiants déclarent avoir triché à un examen majeur. De nombreux travaux ont identifié des fraudes dans le sport (*e.g*., Duggan et Levitt [2002] ; Preston et Szymanski [2003]). Le domaine académique n'est pas épargné par les pratiques non éthiques, allant du *p*-hacking à la fabrication de données (*e.g*., Carrell *et al*. [2008] ; Dufwenberg et Martinsson [2019] dans ce numéro). Dans le vaste domaine des biens de confiance où le consommateur ne peut pas vérifier l'adéquation de la prestation à ses besoins (soins médicaux, réparation automobile, conseil financier, *etc*.), la recherche a montré la haute prévalence des cas de sur-traitement, de sous-traitement et de sur-facturation (*e.g*., Dulleck et Kerchbamer [2006] ; Dulleck *et al*. [2011] ; Balafoutas *et al*. [2013]). A cela s'ajoutent les fraudes aux assurances, la falsification de CV, les fausses évaluations sur internet, *etc*.

La théorie du crime de Gary Becker (Becker [1968]) a guidé pendant des décennies l'analyse économique de ces comportements malhonnêtes, du fait de la beauté de sa formalisation et de ses prédictions simples et claires (voir aussi Ehrlich [1996]). L'application des instruments du calcul coûts-bénéfices a permis aux économistes de contribuer à la compréhension des comportements malhonnêtes aux côtés ou parfois en opposition aux criminologues et spécialistes d'autres sciences sociales (*cf*. le survey de Chalfin et McCrary [2017]). Un agent rationnel évalue le bénéfice attendu d'une action (il-)légale, en temps ou en argent, et les coûts engendrés en termes de probabilité de détection et de sanction en cas de détection. La comparaison des bénéfices nets de l'action illégale et de ceux de l'action légale détermine si l'agent a intérêt à commettre l'infraction. Une implication est que si le risque de détection d'une tricherie est nul, ainsi en est-il du risque de sanction et donc, un agent rationnel devrait tricher sans hésitation dès lors que cela lui procure un bénéfice attendu net. De même, une augmentation de l'écart entre les bénéfices monétaires d'une action malhonnête par rapport à ceux d'une action honnête devrait accroître la probabilité de tricher. A l'opposé, pour décourager la malhonnêteté, il convient d'augmenter la probabilité de détection ou la sévérité des sanctions, les deux instruments étant supposés équivalents si les agents sont neutres au risque.

Les prédictions de la théorie de Becker sont toutefois bousculées par de nombreux travaux expérimentaux récents. Ces recherches, maintes fois répliquées dans des conditions diverses, ont mis en évidence au niveau agrégé que même en l'absence totale de risque de détection et donc de sanction, les individus ne profitent pas de toutes leurs opportunités de s'enrichir en trichant et que la prévalence de la tricherie répond très peu au montant des incitations monétaires. Ces résultats sont troublants pour l'analyse standard et invitent non seulement à incorporer les facteurs moraux dans le processus de décision mais aussi à transformer

fondamentalement la modélisation économique des comportements moraux. C'est ce que je chercherai à montrer dans la première partie de cet article.[1]

La plupart des individus réagissent aux incitations monétaires et préfèrent gagner davantage que moins. Mais pas à n'importe quelles conditions. Quelles sont donc les raisons émotionnelles, normatives, cognitives, qui poussent les individus à renoncer à s'enrichir en trichant ? La deuxième partie de l'article explore le raisonnement moral et montre que face à un dilemme (gagner plus en trichant *vs*. gagner moins en restant intègre), la plupart choisissent de rester honnêtes ou de mentir seulement partiellement. La raison en est une profonde aversion au mensonge, la préservation de l'image de soi et les croyances sur la façon dont les autres perçoivent leurs déclarations. Les nouveaux modèles comportementaux de l'action morale accordent en effet une importance cruciale aux croyances.

Si les individus ont une profonde aversion à la tricherie, alors comment expliquer la prévalence des fraudes ? Si la violation des règles génère un coût moral et heurte les préférences, des stratégies sont mobilisées pour éviter de se confronter aux externalités négatives de ses actes. Par exemple, même un égoïste peut donner de l'argent à un quêteur pour ne pas attenter à son image sociale, mais il peut aussi agir égoïstement tout en évitant ce coût moral en changeant de trottoir. La troisième partie de cet article explore les stratégies d'évitement des coûts moraux des comportements non-éthiques, en particulier l'utilisation de l'incertitude et de récits ou l'évitement de l'information pour accroître sa marge de manœuvre morale. Elle montre que des limitations cognitives sont en fait motivées via la manipulation de ses croyances. Alors que les sections précédentes se centrent sur le mensonge et la tricherie, celle-ci traite de manière plus générale de la moralité, entendue comme la ligne de conduite suivie en présence de dilemmes éthiques quand les choix engendrent des externalités notamment en matière d'honnêteté ou de bienveillance.

LA THÉORIE BECKERIENNE DU CRIME A-T-ELLE VECU ?

De nouveaux outils méthodologiques

Les prédictions de la théorie beckerienne du crime sont simples et élégantes mais elles sont difficiles à tester empiriquement avec les méthodes classiques. Tout d'abord, l'accès aux données est complexe. Les actes répréhensibles sont généralement commis discrètement et les enquêtes directes auprès des individus (par exemple, sur leurs pratiques de parking illégal, de tricherie aux examens, d'évasion fiscale) minimisent les tricheries réelles tant sur leur marge extensive que sur leur marge intensive. Mais même si l'on accéde à des données de qualité, on se heurte à de redoutables problèmes d'estimation en raison de biais d'endogénéité, de simultanéité ou de variables omises. Par exemple, l'estimation de l'impact des effectifs de police sur la criminalité par une corrélation directe entre ces deux variables n'a que peu de valeur car la détermination de ces effectifs résulte en partie du niveau de criminalité.

[1] D'autres revues de littérature sur le mensonge incluent Irlenbusch *et al*. [2015], Jacobsen *et al.* [2017], Abeler *et al*. [2019].

L'économie expérimentale offre des ressources inestimables complémentaires pour appréhender les déterminants des comportements de tricherie. La méthode permet d'introduire de manière contrôlée les variables pertinentes du modèle et seulement ces variables. Grâce à ce degré élevé de contrôle, elle permet de faire varier de manière exogène et contrôlée la valeur des paramètres afin d'identifier des relations de causalité, toutes choses égales par ailleurs, et non seulement des corrélations entre variables, ce qui résout un certain nombre de problèmes économétriques d'endogéneité. Elle permet d'observer les réactions individuelles directes face aux variations des coûts et bénéfices de la tricherie. De plus, la publicité des protocoles utilisés permet leur reproduction dans différents environnements et cultures, ce qui permet non seulement d'identifier des aspects culturels des comportements de fraude mais aussi de soumettre les résultats à des tests de robustesse. Enfin, l'expérimentaliste peut mettre en œuvre une batterie de dispositifs techniques et de procédures pour la collecte de données physiologiques (mesures de la conductance de la peau, rythme cardiaque, dosages hormonaux de testostérone, *e.g.*, Coricelli *et al*. [2010] ; Wibral *et al*. [2012] ; Shalvi et De Dreu [2014]), neurales (par stimulation électrique transcraniale directe, *e.g.*, Maréchal *et al*. [2017], ou par imagerie cérébrale en scanner, *e.g.*, Qu *et al*. [2019]) et cognitives (oculométrie, tests psychologiques, *e.g.*, Wang *et al*. [2010] ; Pitarello *et al*. [2015]) susceptibles d'éclairer les processus engagés lors de la prise de décision morale.

L'économie expérimentale propose un ensemble de jeux standardisés pour étudier l'honnêteté. Ainsi, typiquement dans un jeu d'évasion fiscale, les sujets reçoivent un revenu qu'ils doivent déclarer, connaissant la probabilité d'un contrôle et la pénalité associée en cas de fraude détectée. Il est ainsi possible de tester l'hypothèse de dissuasion (« *deterrence* ») de Becker en manipulant la probabilité de contrôle *vs*. le montant de l'amende, mais aussi de mesurer l'impact d'un report dans le temps de la sanction, de la concentration des contrôles dans le temps, de l'incertitude sur la probabilité de détection, ou encore d'un contrôle à la période précédente sur la conformité à la période suivante.

Pour tester des jeux de communication stratégique à information asymétrique (à la Crawford et Sobel [1982]), un envoyeur informé envoie un message à un receveur qui doit choisir entre deux options qui déterminent les gains des deux joueurs sans connaître à aucun moment les gains respectifs de chaque option (Gneezy [2005]). Il est ainsi possible de tester si le message, et donc la propension à mentir, est sensible à la manipulation des gains relatifs entre les joueurs. Ceci permet d'identifier une variété de mensonges allant du mensonge strictement égoïste, au mensonge Pareto optimal où les deux joueurs retirent avantage du mensonge et au mensonge « blanc » dont seul le receveur bénéficie (Erat et Gneezy [2011]). Ce protocole permet de se rapprocher des déterminants fondamentaux de la prise de décision éthique puisqu'il est de connaissance commune que le receveur ne sera jamais informé des gains relatifs, il n'y a aucun contrôle du message et aucune sanction. Mais ici, préférences sociales et morales peuvent se confondre : si je renonce à un mensonge blanc, est-ce parce que l'utilité du receveur n'entre pas dans ma fonction d'utilité, parce que je crois que le receveur va faire le contraire de ce que je lui recommande, ou que j'ai un coût à mentir infini ?

Les protocoles utilisés pour analyser la propension individuelle à la tricherie se sont encore épurés avec le paradigme du « dé dans la tasse » créé par Fischbacher

et Höllmi [2013] et amendé par Shalvi *et al*. [2011] dans un contexte où le choix du joueur n'impacte que son propre gain. Un dé est placé dans une tasse fermée dont le couvercle est percé d'un trou à travers lequel le joueur peut vérifier de manière privée le résultat du lancement du dé. Ce dispositif fait comprendre aux sujets que l'expérimentaliste ne peut pas identifier un mensonge. Le joueur lance le dé dans la tasse autant de fois qu'il le souhaite mais il doit rapporter le résultat du premier lancé. Son gain dépend du résultat rapporté. Si ce jeu ne permet pas d'identifier la tricherie au niveau individuel, la distribution statistique des réponses de l'ensemble des sujets comparée à la distribution uniforme permet d'estimer un taux de mensonge dans la population. La comparaison de traitements ou la segmentation de l'échantillon (par âge ou genre des sujets) permet d'identifier à quelles variables réagit le taux de tricherie. Cette méthode se décline en de multiples variantes, l'essentiel étant d'utiliser un mécanisme aléatoire dans lequel il y a une correspondance connue entre les gains et les valeurs rapportées. Cela inclue le lancement de pièces de monnaie (*e.g*., Bucciol et Piovesan [2011] ; Abeler *et al*. [2014] ; Cohn *et al.* [2014]), l'ouverture d'une enveloppe remplie de tickets ou de boites sur un écran (*e.g*., Kajackaite et Gneezy [2017] ; Gneezy *et al*. [2018]).

Alternativement, on utilise des tâches à effort réel (par exemple, résoudre des puzzles ou compter des chiffres dans des matrices) où l'on demande aux sujets de déclarer leur performance (*e.g*., Mazar *et al.* [2008] ; Charroin *et al*. [2019]). Quand il est impossible de vérifier la performance réelle, l'identification de la tricherie passe par la comparaison des performances rapportées avec celles d'une groupe de contrôle dont les performances sont vérifiées.

Ces protocoles épurés permettent de tester si comme le prédirait Becker, tous les individus mentent et ce, complètement (par exemple, ils rapportent tous le nombre du dé associé au gain maximum ou la performance maximale). Si tel n'est pas le cas, il est alors possible de caractériser ce qui empêche les joueurs de mentir. Mais si le laboratoire offre des ressources précieuses pour détecter les déterminants des comportements de tricherie, il soulève toutefois une question évidente : quelle est la validité externe des résultats (Levitt et List [2007]) ? La question est d'autant plus légitime que même si des précautions extrêmes sont prises pour éviter le sentiment d'observation, en matière de décision morale les sujets peuvent tout de même se sentir sous surveillance. Deux voies ont été suivies pour rassurer la communauté scientifique. Premièrement, les protocoles ont été sophistiqués pour convaincre les sujets que l'expérimentaliste ne peut pas connaître la vérité, par exemple par les « jeux de l'esprit » (Jiang [2013]). Dans ces jeux, les sujets doivent prédire dans leur tête le résultat du lancement du dé et rapporter uniquement si la réalisation correspond à la prédiction, auquel cas ils remportent un gain.

Deuxièmement, des tests ont mesuré si la tricherie dans les jeux de laboratoire corrèle avec la malhonnêteté dans la vie. Dans Dai *et al.* [2018], nous avons invité des usagers des transports en commun à participer à une expérience dans laquelle nous leur avons administré le protocole du « dé dans la tasse » et d'autres jeux. Une technique nous a permis d'identifier de manière indolore si les participants avaient voyagé en règle avant de participer à l'étude. Les résultats révèlent que les resquilleurs dans les transports mentent significativement plus dans les jeux de laboratoire. Ainsi, au niveau agrégé ces simples jeux expérimentaux permettent de prédire dans une certaine mesure la moralité dans la vraie vie courante. Des résultats

convergents ont corrélé une distribution non uniforme des valeurs rapportées dans les jeux de dés avec le fait pour des vendeurs en Inde de couper le lait (Kröll et Rustagi [2016]), être un prisonnier condamné à une plus forte peine (Cohn *et al.* [2015]), être un étudiant perturbateur à l'école (Cohn et Maréchal [2019]), ne pas rendre un trop-perçu à la fin d'une expérience (Potters et Sloop [2016]).

Des résultats troublants pour la théorie Beckerienne

Ces travaux ont délivré des conclusions troublantes pour la théorie du crime rationnel. L'impressionnante méta-analyse réalisée par Abeler *et al*. [2019] à partir de 90 études utilisant le paradigme du « dé dans la tasse » ou ses variantes (donc sans interactions stratégiques), impliquant plus de 44000 sujets constituant autant d'observations indépendantes, vient le confirmer (voir ses détails à www.preferencesfortruthtelling.com). Trois résultats sont particulièrement forts :

- Tous les individus ne mentent pas dans des contextes où le mensonge permet de maximiser ses gains, sans risque de détection et sans conséquences négatives pour autrui (si ce n'est l'expérimentaliste).
- La plupart des individus ne mentent pas complètement, mais seulement partiellement. Dans le jeu initial Fischbacher et Föllmi-Heusi [2013] estiment qu'environ 20% des joueurs mentent complètement, 39% rapportent honnêtement et les autres joueurs mentent partiellement.
- Une augmentation du montant des incitations à mentir n'a pas d'effet significatif sur la distribution des réponses des sujets.

La méta-analyse d'Abeler *et al*. [2019] confirme les résultats initiaux : toutes les valeurs possibles sont rapportées par les sujets, pas seulement celles qui maximisent les gains mais aussi celles qui les minimisent. Donc, même sans risque de détection et en présence d'incitations à tricher, tous ne mentent pas (résultat aussi observé par Abeler *et al*. [2014] dans une expérience par téléphone). Lorsque la distribution théorique est uniforme, la probabilité de rapporter un montant donné augmente faiblement avec sa valeur. Ceci suggère une faible sensitivité individuelle à l'incitation monétaire. Abeler *et al*. estiment de manière agrégée que les individus renoncent environ aux trois quarts des gains potentiels du mensonge, augmenter de 500 fois le montant des enjeux ne conduit pas à des résultats différents, répéter la tâche des dizaines de fois ne génère pas d'apprentissage. Une expérience de terrain récente conduite dans 40 pays et dans laquelle l'expérimentaliste prétend avoir trouvé un porte-monnaie dans la rue et demande à un sujet de le renvoyer à son propriétaire bouscule également la vision beckerienne : plus la somme contenue dans le porte-monnaie est élevée, plus le taux de retour est élevé (Cohn *et al*. [2019]), phénomène déjà observé dans une expérience où les clients d'un restaurant recevaient un trop-perçu (Azar *et al*. [2013]).

La question du mensonge partiel est intrigante. Fischbacher et Föllmi-Heuser [2013] ont montré que le mode de la valeur rapportée dans le jeu de dé n'est pas la valeur maximale mais la seconde valeur (4 dans leur protocole). Ceci est confirmé par Abeler *et al*. [2019]. Ce résultat laisse entendre que certains individus sont tentés de mentir pour gagner davantage mais n'osent pas mentir complètement même en l'absence de risque. Au total, ces recherches nombreuses montrent que la

plupart des individus éprouvent de l'aversion au mensonge dans un contexte dépouillé de toute institution et interactions sociales.

COÛTS MORAUX, AVERSION A LA TRICHERIE ET REPUTATION

Avant de revenir sur les mécanismes qui génèrent de l'aversion morale au mensonge, arrêtons-nous sur le rôle des préférences et des biais comportementaux qui perturbent l'appréhension des gains et coûts monétaires associés à la tricherie.

Aversion à la perte et biais d'appréhension des gains monétaires de la tricherie

Si les travaux récents ont montré une élasticité faible du mensonge aux gains espérés, cela ne signifie pas pour autant qu'aucune dimension des incitations monétaires n'influe sur ces comportements. D'une part, la nature des incitations conditionne la propension à mentir. Par rapport à des rémunérations fixes, les modes de rémunération à la performance individuelle poussent à la tricherie (*e.g.*, Cadsby *et al*. [2010] ; Belot et Schröder [2013] ; Gravert [2013]). Il en est de même des rémunérations compétitives et des rémunérations de groupe (*e.g.*, Jacob and Levitt [2003] ; Conrads *et al*. [2013, 2014] ; Danilov *et al.* [2013] ; Gill *et al*. [2013] ; Buser et Dreber [2015], s'ajoutant à la littérature sur le sabotage dans les tournois). Ceci s'explique en partie parce que les modes compétitifs attirent les individus qui ont un moindre coût de tricherie (Faravelli *et al*. [2015]). Cela ne signifie pas pour autant que les rémunérations fixes sont un remède contre la tricherie. Nous avons montré dans Charness *et al*. [2014] le rôle de la comparaison des performances relatives et de la préoccupation des individus pour leur rang. Le goût pour le rang ou la joie de gagner compense chez certains le coût moral du mensonge.

A côté de la nature des incitations monétaires, la probabilité des gains influence la propension à tricher. Dans Garbarino *et al*. [2019], nous proposons un modèle de mensonge où les individus averses à la perte ajustent leur comportement moral aux variations de la probabilité d'un gain. Le résultat espéré d'une loterie constitue un point de référence. Ensuite, les individus observent le résultat du tirage au sort qui est plus ou moins distant du point de référence. Mentir sur le résultat observé est une façon de se rapprocher de ce point. Si les individus souffrent davantage de pertes qu'ils ne se réjouissent de gains équivalents (Kahneman et Tversky, [1979]), ils gagnent d'autant plus d'utilité en mentant que le résultat obtenu est inférieur et distant du gain attendu. Une probabilité accrue de gain élevé augmente leur espérance de gain et si le mauvais résultat se réalise, leur probabilité de mentir augmente pour éviter une perte plus grande. Le test de cette théorie sur le même panel que celui utilisé par Abeler *et al*. [2019] et lors de nouvelles expériences sur MTurk soutient la relation négative entre la probabilité du bon résultat et le taux estimé de mensonge (pour une méthode de calcul du taux de mensonge quand la vérité est inconnue de l'expérimentaliste, voir Garbarino *et al*. [2018]). D'autres études ont également trouvé un effet de l'aversion à la perte sur le mensonge à partir d'une manipulation du montant de la dotation initiale pour créer un effet de cadrage. Les gens mentent plus pour éviter une perte que pour gagner (Cameron et Miller

[2013]; Grolleau *et al*. [2016]; Schindler et Pfattheicher [2016]; Balasubramanian *et al.* [2017]) ; seuls Charness *et al*. [2019] trouvent un effet différent.

A côté de l'aversion à la perte, d'autres préférences individuelles expliquent que les individus réagissent de manière non linéaire au montant des incitations. On pense bien entendu aux attitudes face au risque mais aussi aux préférences temporelles s'il existe un risque de détection (lesquelles ne jouent aucun rôle dans le jeu de dé précédent). Par exemple, Lee et McCrary [2019] montrent comment les prédictions de Becker doivent être amendées si l'on intégre l'hétérogénéité des préférences temporelles des agents. Par exemple, un allongement de la peine de prison a un effet plus limité que prédit si les contrevenants sont impatients. Aussi, Ruffle et Tobol [2014] montrent qu'en raison de l'escompte du temps, la probabilité de mentir augmente lorsque la récompense est plus proche dans le temps.

Ces analyses mettent en avant le rôle des préférences et des biais cognitifs dans l'appréhension des gains et des pertes monétaires d'une tricherie. Un tout autre type d'analyses, examinées dans les sections suivantes, met en avant l'aversion à la tricherie et les coûts moraux, les normes sociales, les émotions et les croyances.

Sensibilité aux coûts moraux du mensonge et aux normes injonctives

Nous apprenons dès l'enfance de nos parents qu'il est « mal de mentir » (Houser *et al*. [2016]). Ne pas dire la vérité est coûteux moralement et même si les enfants développent des préférences sociales en grandissant, ils ne deviennent pas davantage prêts à mentir, même pour aider les autres, car ils apprennent une éthique morale de la vérité en même temps que les préférences sociales (Maggian et Villeval [2015]). La socialisation augmente le coût moral d'un non-respect des règles et normes injonctives, de l'impératif Kantien de vérité. On ment moins si l'on pense que la plupart des gens pensent que mentir est inapproprié. Pourtant, le coût de déviation de la norme injonctive est sensible aux conditions environnementales et sociales en vigueur.

Sur le premier point, des travaux en psychologie ont montré que le mensonge est favorisé par l'épuisement cognitif (*e.g*., Mead *et al.* [2009]; Gino *et al*. [2011] ; Kouchaki et Smith [2014]), la faim (Polizzi di Sorrentino *et al*. [2019]), la pression du temps (Shalvi *et al*. [2012]). Des préférences économiques standard dans un modèle économique standard prédiraient au contraire que l'individu est davantage capable de tricher rationnellement quand il est en possession de tous ses moyens. La présence de signaux moraux conditionne aussi la tentation de mentir si les individus s'inquiètent de leur réputation (voir la section suivante). Par exemple, réciter les 10 commandements (Mazar [2008]), placer une signature sur l'honneur en haut d'une déclaration de sinistre (Shu *et al*. [2012]), signer un serment avant de participer à l'expérience (Jacquemet *et al.* [2019]), rappeler qu'il est moral de payer son dû (Pruckner et Sausgruber [2013]) peuvent réduire la tricherie en augmentant les coûts moraux associés. Toutefois, l'utilisation de *nudges* dont le mécanisme repose sur les normes sociales et vise à modifier les croyances sociales est parfois contre-productif (Bicchieri et Dimant [2019]). Quand normes injonctives et normes empiriques (voir ci-dessous) se contredisent, diffuser des messages normatifs s'avère inefficace (Bicchieri et Xiao [2009]).

Les conditions sociales affectent aussi les coûts moraux. La propension à mentir est inférieure quand il existe une victime, comme dans les jeux de communication stratégique, et plus le coût pour la victime est important (Gneezy [2005]). Inversement, elle peut s'accroître si le mensonge bénéficie à autrui (« *white lies* ») ou à la fois à soi-même et à autrui (« mensonge Pareto-optimal ») (Erat et Gneezy [2012]). Mais même dans ce contexte, tous ne mentent pas, ce qui tend à attester d'une aversion au mensonge (un coût infini de mentir). Par ailleurs, l'identité de groupe module l'attention portée à la victime. Elle conduit à tricher davantage pour favoriser quelqu'un appartenant à son propre groupe social (*e.g*., Gino *et al*. [2009] ; Fosgaard *et al*. [2013] ; Cadsby *et al*. [2016]), notamment sous l'influence de l'oxytocine (Shalvi et de Dreu [2014]). De même en est-il pour se venger de personnes d'un autre groupe que le sien. Dans Banerjee *et al*. [2018] notre expérience de terrain en Inde montre qu'après avoir perdu un tournoi, les personnes de la caste élevée mentent davantage au jeu de dé pour réduire le gain d'un joueur de la caste inférieure. Il semble moralement plus acceptable de dévier de la ligne de conduite éthique quand c'est pour le bénéfice d'un allié ou au détriment de quelqu'un qui n'appartient pas à son cercle. Bénistant et Villeval [2019] montrent toutefois que l'identité de groupe a peu de poids dans un contexte de pure compétition où la volonté de gagner est la plus forte.

Les coûts moraux de la tricherie sont aussi influencés par les institutions qui gouvernent nos sociétés. Cette question renvoie à un débat fondamental sur la façon dont ces institutions affectent nos valeurs morales. Pour certains, les institutions marchandes érodent les valeurs morales en prônant l'individualisme et les valeurs matérielles (Sandel [2012] ; Falk et Szech [2013]). Pour d'autres, il n'est pas de tendance universelle (Bartling *et al*. [2015] ; Bartling et Ozdemir [2017]). Les institutions informelles, en particulier les normes empiriques (ou « descriptives »), affectent en effet les coûts du mensonge. Les actions sont influencées non seulement par ce que les individus pensent que les autres considèrent comme approprié, mais aussi par la façon dont ils imaginent ce que les autres font ou feraient dans la même situation (*i.e*., les normes descriptives, Bicchieri [2006]). Ils mentent moins s'ils pensent que la plupart des personnes ne mentent pas et que, conditionnellement à ces croyances, ils ont une préférence pour se conduire honnêtement. Mais si les individus ont des préférences conditionnelles, dans le cas où leurs croyances changent ils peuvent cesser de suivre la norme (Bicchieri *et al*. [2019a] ; voir aussi Rauhut [2013]). Ainsi, les effets de pairs favorisent la diffusion du mauvais exemple au sein des groupes (*e.g*., Carrell *et al*. [2008]; Gino *et al*. [2009]; Lefebvre *et al.* [2015] ; pour un résultat plus mitigé sur l'évasion fiscale, voir Fortin *et al*. [2007]). Certaines cultures d'entreprise favorisent aussi la fraude. Par exemple, Cohn *et al*. [2014] montrent que lorsque l'on active au préalable leur identité professionnelle, les banquiers ont tendance à mentir davantage dans un jeu de lancer de pièces alors qu'ils sont plus honnêtes que la moyenne quand leur identité professionnelle n'est pas activée. Cela s'explique sans doute parce que cette manipulation réactive les valeurs matérialistes liées à leur profession.

Enfin, des comparaisons internationales montrent qu'il existe une corrélation positive entre la faiblesse des institutions politiques et juridiques d'un pays et le taux de tricherie dans le jeu de dé présenté dans la section précédente (Gächter et Schultz [2016]) ou encore que l'érosion des normes morales par le marché est

sensible aux conditions culturelles en vigueur (Bartling *et al*. [2015]). Là encore, le comportement moyen des membres du groupe module la force des valeurs morales.

Les explications suivantes insistent soit sur la culpabilité, soit sur les croyances sociales mais ils relèvent de la même approche où l'utilité des agents dépend des croyances.

Emotions et culpabilité

La plupart des individus cherchent à conserver une image positive d'eux-mêmes en ne décevant pas les autres. La théorie des jeux psychologiques, introduite par Geanakoplos *et al*. [1989] et développée par Battigalli et Dufwenberg [2009], est un outil théorique précieux pour comprendre la nature de ces coûts moraux en modélisant des émotions comme l'aversion à la culpabilité.[2] Ainsi, à l'aide d'un jeu de confiance Charness et Dufwenberg [2006] ont étudié l'impact de la communication non contraignante en amont du jeu sur la confiance. Ils ont ainsi montré qu'une partie significative des individus ne veulent pas trahir leur promesse d'un choix coopératif pour influencer l'action d'un autre joueur à leur avantage. La communication, sous forme de promesse, soutient la confiance et la coopération parce que personne n'aime décevoir les attentes d'autrui.[3] Le mécanisme sous-jacent est le fait que les promesses modifient ces croyances.

L'aversion à la culpabilité (le fait que les individus se sentent coupables de decevoir ce qu'ils pensent être les attentes des autres) contribue aussi à expliquer les attitudes face à la tricherie. Par exemple, dans un jeu de communication stratégique à information asymétrique à la Gneezy [2005], Lundquist *et al*. [2009] montrent que l'aversion au mensonge augmente avec la taille du mensonge et la force des promesses. Dans Attanasi *et al*. [2019], nous avons étudié théoriquement et expérimentalement l'aversion à la culpabilité dans le contexte de détournements de fonds, à l'aide d'un jeu à information incomplète avec aversion à la culpabilité dépendant du rôle. Dans un mini-jeu de détournement à trois joueurs, le donateur peut envoyer de l'argent à un bénéficiaire via un intermédiaire. Ce dernier prend une décision pour chaque croyance possible du bénéficiaire sur le montant reçu, et, dans un autre traitement, pour chaque croyance possible du donateur sur la part du don transférée. Le détournement réduit l'utilité du donateur qui se préoccupe du bien-être du bénéficiaire et celle du bénéficiaire qui est spolié. Nous confirmons l'aversion à la culpabilité comme motivation dépendant des croyances : environ un quart des intermédiaires ajustent leur décision en fonction des croyances des autres joueurs dans le sens attendu. De façon intéressante, l'aversion à la culpabilité a la même intensité vis-à-vis du donateur (bien que son gain ne soit pas affecté par le

[2] La théorie des jeux psychologiques se distingue de la théorie des jeux classique par la prise en compte des croyances de premier et de second ordre dans la fonction d'utilité des individus. Leurs préférences dépendent de leurs croyances sur les croyances des autres joueurs. Pour une application à l'étude de l'altruisme, voir Attanasi *et al*. [2019] dans ce numéro.

[3] Au contraire, Vanberg [2008] soutient que si les individus tiennent leurs promesses c'est à cause d'une préférence intrinsèque pour tenir leurs engagements plus qu'en raison d'une volonté de ne pas décevoir les attentes d'autrui. Voir aussi López-Pérez et Spiegelman [2013] qui identifient une pure aversion au mensonge indépendante du souci de se conformer aux attentes des autres.

détournement) et vis-à-vis du bénéficiaire. Des approches plus générales s'appuyant sur les croyances de second ordre sont décrites à la section suivante.

D'autres émotions sociales proches de la culpabilité ont été également décrites dans la littérature, sans faire nécessairement référence à la théorie des jeux psychologiques. Ainsi, dans Coricelli *et al*. [2010], nous avons manipulé dans un jeu d'évasion fiscale l'exposition publique des fraudeurs en cas de fraude détectée et mesuré l'intensité émotionnelle grâce à des mesures électro-physiologiques. Les résultats ont mis en évidence les sentiments de honte éprouvés en cas de détection mais également un rôle des émotions anticipatrices au moment de la décision de déclaration (voir aussi Coricelli *et al*. [2012] ; B*ø et al*. [2015]).

De nouveaux modèles intégrant croyances et réputation

Ainsi que modélisé notamment dans Bénabou et Tirole [2002], le désir de se voir sous une lumière positive est une préoccupation essentielle pour la plupart des êtres humains, lesquels cherchent à préserver à la fois leur image de soi et leur image sociale (voir Qu *et al*. [2019] pour un test neuro-économique de ce modèle). Concernant l'image de soi, Mazar *et al.* [2008] et Fischbacher et Föllmi-Heusi [2013] expliquent le mensonge partiel par le fait que les individus cherchent avant tout à conserver une bonne image d'eux-mêmes (« *self-image maintenance* »), ce qui conduit à un coût marginal du mensonge croissant avec la taille du mensonge. Le coût intrinsèque du mensonge est également supérieur quand on ment sur sa performance (donc sur son habilité, une caractéristique qui définit son être) que sur sa chance dans un tirage au sort, plus impersonnelle (Kajackaite [2016]). Concernant les préoccupations d'image (ou d'identité) sociale, celles-ci dépendent des croyances que l'individu forme sur les soupçons des autres à son égard et peuvent aussi expliquer le mensonge partiel (voir aussi Tergiman et Villeval [2019] qui montrent comment la réputation change aussi la nature des mensonges dans un jeu à information asymétrique). Plusieurs modèles théoriques inspirés de la théorie des jeux psychologiques ont été produits récemment pour expliquer l'importance de combiner ces deux dimensions.

S'intéressant au paradigme du jeu de dé, Dufwenberg et Dufwenberg [2018] introduisent l'hypothèse que les individus souffrent d'une aversion à être perçus comme menteurs lorsqu'ils rapportent leur résultat à une autorité (par exemple, l'expérimentaliste). Le modèle propose une fonction d'utilité dépendant des croyances et explique pourquoi rapporter le gain maximum ne correspond pas à la maximisation de l'utilité. Si l'individu pense qu'en rapportant le gain maximum l'autorité va inférer qu'il a menti, la désutilité d'être identifié comme tricheur potentiel réduit son incitation à mentir. Le modèle peut substituer l'audience par la propre conscience de l'individu. Cette désutilité est proportionnelle à la taille du mensonge supposée par l'audience. Elle s'ajoute aux coûts directs de mensonge qui peuvent être linéaires ou convexes avec la taille du mensonge (Mazar *et al.* [2008] ; Fischbacher et Föllmi-Heusi [2013]) ou bien fixes (voir Kartik [2009] et Kajakaite et Gneezy [2017]). Cette modélisation des préférences peut expliquer la coexistence de comportements de maximisation des gains, de vérité, et de mensonges partiels sans avoir à supposer l'hétérogénéité de types de joueurs.

Les modèles de Gneezy *et al.* [2018], Abeler *et al*. [2019] et Khalmetski et Sliwka [2019] prennent aussi en compte l'utilité d'être perçu comme une personne honnête comme élément de définition de l'identité. Une fonction d'utilité additive inclut les coûts réputationnels du mensonge aux côtés d'un coût intrinsèque à mentir et des gains monétaires du mensonge. Dans ces modèles qui utilisent aussi les outils de la théorie des jeux psychologiques, cette désutilité est associée à la probabilité que l'audience attribue au sujet d'être un menteur, pas au montant supposé du mensonge à la différence de Dufwenberg et Dufwenberg [2018].

Gneezy *et al.* [2018] montrent qu'à l'équilibre, il existe une valeur seuil du résultat réel au-delà duquel les individus ne mentent jamais et les individus disent des mensonges partiels s'ils accordent suffisamment d'importance à leur réputation. En manipulant la distribution des gains, l'observabilité de la vérité et la probabilité du gain maximum dans la loterie, ils confirment que les individus sont sensibles à leur réputation et que la taille du mensonge affecte directement le coût intrinsèque de mentir. Ils ont ainsi une vision multidimensionnelle du coût du mensonge qui dépend de sa taille. Par exemple chez Lundquist *et al.* [2018] la taille dépend uniquement du résultat, *i.e*., la distance à la vérité ; ici elle dépend aussi du gain du mensonge et de la probabilité du bon résultat. Si l'expérimentaliste peut vérifier *ex post* l'existence d'un mensonge au niveau individuel, certains individus n'ont pas de réticence à mentir maximalement. En revanche, si l'expérimentaliste ne peut pas vérifier la véracité de la déclaration au niveau individuel, il existe davantage de mensonges partiels et la fréquence de ces mensonges partiels augmente en relation inverse avec la probabilité *ex ante* du meilleur résultat possible au tirage au sort. Ce lien négatif entre probabilité du résultat maximum et mensonge partiel est similaire à ce qui a été identifié par Garbarino *et al.* [2019], mais le mécanisme est la volonté de s'enrichir tout en préservant sa réputation et non pas l'aversion à la perte.

De leur côté, Abeler *et al*. [2019] considèrent trois types d'explication du fait que les individus ne mentent pas complètement dans des tâches où le mensonge n'impacte les gains de personne d'autre : les coûts directs du mensonge, l'impact des normes (conformité, aversion à l'inégalité) et la réputation (réputation d'honnêteté ou aversion à la culpabilité). Ils montrent que les données de leur méta-analyse sont compatibles avec toutes les catégories de modèles. Ils développent alors un nouveau modèle dans lequel l'utilité augmente avec le gain monétaire et qui intègre des préoccupations d'image de soi (coût moral) et d'image sociale (croyances de l'audience sur l'honnêteté du montant rapporté). Et ils conçoivent de nouvelles expériences pour discriminer entre les modèles en manipulant la distribution des états réels, les croyances sur la norme empirique (sans effet), l'observabilité du résultat réel (produisant un fort effet) [4]. La conclusion est que seuls les modèles qui combinent une préférence pour l'honnêteté [5] *et* une préférence pour être perçu comme quelqu'un d'honnête ne sont pas réfutés.

[4] L'effet de l'observabilité n'est toutefois pas systématique. Par exemple, Fischbacher et Föllmi-Heusi [2013] ne trouvent pas de différence de comportement quand ils utilisent une procédure en double-aveugle pour le jeu de dé. Van de Ven et Villeval [2015] ne trouvent pas d'effet de la levée de l'anonymat dans un jeu d'envoyeur-receveur, possiblement parce que dans ce jeu, les gains dépendent de la décision du receveur, ce qui dilue la responsabilité de l'envoyeur.

[5] Dans le passé, l'aversion intrinsèque au mensonge a déjà fait l'objet de travaux théoriques (*e.g*., Kartik *et al*. [2007] ; Kartik *et al*. [2014]).

LES STRATEGIES DE MINIMISATION DES COÛTS MORAUX ET REPUTATIONNELS

Les individus ont une préférence intrinsèque pour l'honnêteté et une préférence à être perçus comme des êtres moraux. Il n'en reste pas moins que de nombreuses fraudes et malhonnêtetés sont constatées dans des contextes variés. Une explication réside dans les institutions qui entourent la prise de décision (les structures de marché et les institutions politiques, notamment) et qui affectent le poids relatif des motivations monétaires et des motivations morales dans la fonction d'utilité des agents. Ceci mériterait un article entier et nous n'aborderons pas ce thème ici. Une autre explication se trouve dans les stratégies déployées par les individus, consciemment ou pas, pour bénéficier des avantages monétaires d'un comportement égoïste tout en continuant à se percevoir comme des êtres moraux. Nous évoquerons dans cette dernière section trois ensembles de stratégies qui illustrent des formes d'« hypocrisie morale » (*e.g.*, Batson *et al*. [1999] ; Rustichini et Villeval [2014]) dans les comportements face aux dilemmes moraux : les croyances motivées et la recherche d'incertitude ; le désengagement moral et l'auto-justification ; l'utilisation de récits narratifs dans la recherche d'excuses.

Recherche d'incertitude et distorsion motivée des croyances

L'*incertitude* et l'*ignorance stratégique* réduisent les coûts moraux et la responsabilité de ceux qui commettent des choix non éthiques. Rester ignorant des externalités négatives produites par ses actes évite d'en subir le coût moral. Gino *et al*. [2016] parlent de « sujets bayésiens motivés », Bénabou [2015] de « croyances motivées » et Exley et Kessler [2018] de « limitations cognitives motivées » pour décrire le traitement biaisé de l'information et la distorsion de l'information.

La distorsion des croyances peut excuser un comportement non éthique ou égoïste. Dana *et al*. [2007] ont identifié comment les individus utilisent l'incertitude pour se créer une marge de manœuvre morale à l'aide de jeux de dictateur dans lesquels le lien entre les choix et leurs conséquences pour autrui est incertain soit pour le dictateur, soit pour le receveur. Dans cet environnement, les joueurs se comportent de manière plus égoïste que lorsqu'ils sont informés parce que l'ignorance protège leur image (voir également Grossman [2014] ; Kajackaite [2015] ; Grossman et van der Weele [2017]). Utilisant des données naturelles suédoises, Freddi [2019]) montre que les personnes vivant dans des zones où la présence de réfugiés a augmenté lisent moins d'articles de presse parlant des demandeurs d'asile, surtout si ces articles témoignent d'empathie pour les réfugiés. Un autre avantage de l'ignorance est que les dictateurs (volontairement) ignorants sont moins punis par des tiers que lorsque la même décision est prise en connaissance de cause ; toutefois, rechercher l'ignorance suscite des sanctions (Bartling *et al*. [2014]). S'intéressant au rôle du conseil en présence de dilemmes éthiques, dans Shalvi *et al*. [2019] nous identifions l'existence d'une offre

d'ignorance qui répond à cette demande d'ignorance et nous montrons qu'il existe un marché de l'ignorance qui se révèle efficient dans l'appariement de l'offre et de la demande d'ignorance. Coffman et Gottard Real [2019] montrent que lorsque l'action non éthique résulte d'un conseil, les victimes punissent moins, ce qui se traduit par plus d'égoïsme lorsqu'on peut utiliser des conseillers. Forcer les décideurs à se confronter à l'information réduit l'égoïsme (Lind *et al.* [2016]).

Étudiant les comportements de don, Exley [2015] montre comment le risque est utilisé comme excuse pour ne pas donner. Face aux mêmes loteries, les sujets sont davantage averses au risque quand un don est en jeu. Une forme supérieure d'incertitude est l'ambiguïté où l'on ne connaît pas la probabilité d'un évènement aléatoire. A l'aide de jeux de dictateur, Haisley et Weber [2010] ont identifié que l'ambiguïté est biaisée en sa faveur lorsque le gain du receveur est incertain, ce qui conduit à une moindre générosité en situation d'ambiguïté qu'en situation de risque Cettolin *et al*. [2017] ne trouvent toutefois pas de différence. Étendant Exley [2015] à l'ambiguïté, Garcia *et al*. [2019] ne trouvent pas non plus de différence d'impact entre ambiguïté et risque : toute forme d'incertitude est suffisante pour servir d'excuse quand ses décisions créent des externalités négatives.

Une autre forme d'ignorance et d'évitement des informations gênantes est l'*aveuglement moral*. Les zones aveugles en matière éthique *(« ethical blind spots »)* permettent de se comporter égoïstement sans en subir le coût moral. Utilisant l'oculométrie, Pitarello *et al*. [2015] ont ainsi montré que les individus qui mentent tendent à éviter de regarder les informations qui rendraient leur mensonge plus difficile et fixent davantage leur regard sur l'information tentante.

La *mémoire sélective* constitue un autre mécanisme mobilisable pour créer de *l'incertitude ou de l'aveuglement sur ses actions passées*, en particulier pour restaurer la cohérence entre image de soi et réalité de ses actions lorsqu'elles menacent cette image. Le temps fournit une marge de manœuvre pour déformer ses souvenirs, voire en créer de faux afin de réconcilier le soi présent et le soi futur (Bazerman *et al*. [1998] ; Benabou et Tirole [2002] ; Moore [2016]). Il existe deux modes principaux de motivation de sa mémoire. Le premier est le rappel sélectif de ses souvenirs, c'est-à-dire que la probabilité de se remémorer un acte désirable est supérieure à celle de se remémorer un acte indésirable d'un point de vue éthique (*e.g*., Carrillo et Mariotti [2000] ; Mullainathan [2002]). Le second mode est le biais des erreurs de mémoire et il se définit par la direction et la magnitude des erreurs. Par exemple, Chew *et al*. [2018] ont mis en évidence des phénomènes d'illusions positives ou d'affabulation. La plupart des expériences économiques sur la mémoire mettent en évidence l'asymétrie des souvenirs des individus selon qu'ils reçoivent une bonne ou une mauvaise nouvelle à propos de leur performance dans un test d'intelligence (*e.g*., Chew *et al*. [2018] ; Zimmermann [2019]).

Peu d'expériences économiques ont exploré la mémoire motivée dans le domaine éthique contrairement à la psychologie (*e.g.,* Shu *et al*. [2011] sur la mémoire des mensonges). Saucet et Villeval [2019] ont exposé des sujets à un jeu de dictateur binaire répété offrant des choix plus ou moins inégalitaires entre une option égoïste et une option altruiste. Les dictateurs ont eu ensuite à se rappeler le gain attribué au joueur passif dans chaque décision contre rémunération. Les résultats confirment la sélectivité des souvenirs. Le taux d'erreurs de mémoire est

inférieur quand les sujets ont pris la décision altruiste plutôt que la décision égoïste. Un traitement dans lequel les choix sont faits par le programme informatique montre que la mémoire sélective est indissociable de la responsabilité. En effet, lorsque le choix est fait par le programme, le taux d'erreurs de mémoire ne change pas entre choix égoïstes et altruistes. En revanche, la direction et la magnitude des erreurs de mémoire ne diffèrent pas selon la nature des choix. En d'autres termes, les individus ont une mémoire sélective qui leur fait supprimer l'information qui menace leur image, mais ils ne montrent pas de biais de mémoire motivés. Carlson *et al*. [2018] montrent que la qualité de la mémoire est inférieure lorsque la décision de l'individu s'écarte de sa vision de la justice. Enfin, les individus se remémorent moins bien leurs choix non éthiques, alors que leurs souvenirs ne diffèrent pas entre les actions éthiques et non-éthiques des autres (Kourachi et Gino [2016]).

Désengagement moral et auto-justification

Les psychologues ont montré que les individus mentent d'autant plus qu'ils peuvent se trouver une justification. Par exemple, dans le jeu de dé les joueurs peuvent lancer le dé plusieurs fois mais la règle est de déclarer le résultat du premier lancé. Demander aux joueurs de ne lancer le dé qu'une seule fois conduit à moins de mensonges (Shalvi *et al*. [2011]). En effet, lancer le dé plusieurs fois donne plus de chances au joueur d'observer effectivement le chiffre qui maximise le gain et donc de se sentir moins en faute s'il rapporte ce chiffre (voir aussi Shalvi *et al*. [2015]). S'il est facile de se trouver des excuses, en revanche le désengagement moral ne s'applique pas au jugement sur la conduite des autres (*e.g*., Shu *et al*. [2011]).

L'auto-justification peut s'appuyer également sur la *gestion dynamique des comportements moraux*. Les psychologues ont introduit la notion d' « *équilibre moral* » pour rendre compte du fait que les individus ont tendance à contrebalancer leurs actions au cours du temps. Selon Nisan [1991], les individus évaluent la distance entre leur état moral actuel et leurs normes personnelles. Ils ne cherchent pas à avoir toujours un comportement éthique mais ils ont un seuil minimal à ne pas franchir et ils se comportent de manière à conserver en moyenne leur équilibre. Ainsi, après s'être particulièrement bien comporté, un individu relâche son comportement moral (« *moral licencing* ») (*e.g*., Merritt *et al*. [2010]); inversement, après avoir commis une action amorale, l'individu cherche à se rattraper par un comportement éthique (« *moral cleansing* ») (voir aussi Monin et Jordan [2009], et pour une revue méta-analytique, Blanken *et al*. [2014]). De même, les individus mentent plus quand ils viennent d'être traités de manière malveillante par un autre joueur (Houser *et al*. [2012]), après avoir permis à une association charitable de gagner davantage (Ploner et Regner [2013]) ou quand ils savent l'opportunité d'une bonne action dans le futur proche (Cojoc et Stoian [2014]). Gneezy *et al*. [2014] reproduisent ces phénomènes qu'ils interprètent en termes de « *comptabilité de la conscience* » : après avoir éprouvé un sentiment de culpabilité, l'individu cherche à se racheter et à blanchir ses actes.

Un autre ensemble de stratégies de désengagement consiste à se cacher derrière les autres.

Une façon d'auto-justifier ses comportements non éthiques est de se référer au comportement moyen, c'est-à-dire à la norme du groupe. L'observabilité du

comportement moyen des autres conduit à une érosion de la conformité aux normes morales dans les choix pro-sociaux parce que les exemples de violations de la norme se diffusent – certes, de façon plus nuancée quand les autres sont proches (*e.g*., Bicchieri *et al*. [2019a]). Les individus peuvent aller plus loin en *déformant leurs croyances sur la norme de façon motivée*. Dans une expérience à deux étapes, Bicchieri *et al*. [2019b] montrent que les individus déforment leurs croyances sur la norme en vigueur dans la première étape quand ils sont informés qu'ils auront l'opportunité de mentir dans un jeu de dé en deuxième étape, par rapport à ceux qui n'en sont pas informés. La structure d'information permet aux menteurs de se persuader que le mensonge est répandu mais n'affecte pas la norme injonctive. Autrement dit, ils ont autant conscience que le mensonge est désapprouvé, mais ils se dédouanent de leur immoralité en se persuadant que mentir est la norme. C'est donc l'incertitude sur la norme empirique (ce que les autres font) qui permet le flou moral et la distorsion des croyances à son avantage et non pas l'incertitude sur la norme injonctive (ce que les autres approuvent).

Dans la recherche précédente, les individus manipulent leurs croyances normatives mais une autre façon de s'abriter derrière la norme pour réduire son coût moral à mentir est de choisir son groupe social de référence. Des travaux théoriques ont montré que les processus d'appariement contribuent à l'évolution des préférences morales dans une société (*e.g*., Alger et Weibull [2013]). Charroin *et al*. [2019] illustrent ce phénomène à travers l'étude de l'*homophilie* dans les comportements malhonnêtes lorsqu'il n'y a pas de complémentarités productives. Les participants peuvent ou non mentir sur leur performance selon le mode d'évaluation choisi. Au milieu de l'expérience ils ont l'opportunité de choisir les participants à des sessions passées sur les actions desquels ils recevront de l'information dans la partie suivante, sur la base de leur choix de mode d'évaluation. Nous voyons que les menteurs cherchent à se regrouper avec des menteurs potentiels alors que les joueurs honnêtes ne sont pas homophiles. Choisir un groupe de référence qui semble avoir une faible coût moral à mentir allège son propre coût moral parce que le comportement moyen donne la norme. Dans une expérience où les sujets choisissent de conserver le même partenaire ou d'en changer et où les efforts sont joints, Gross *et al*. [2018] montrent comment les individus malhonnêtes recherchent des partenaires du même type qu'eux. Ils mettent aussi en évidence que les individus honnêtes qui ne mentent pas eux-mêmes cherchent à conserver un partenaire malhonnête pour en tirer avantage, produisant une sorte de *free-riding* éthique.

Une stratégie alternative de désengagement consiste à *déléguer à d'autres* les choix impliquant des dilemmes moraux pour diluer sa responsabilité et libérer sa conscience. Grâce à un jeu de dictateur, Hamman *et al.* [2010] montrent que la délégation de la décision conduit à des choix plus égoïstes et diffuse la responsabilité. Avec la délégation, les individus repoussent le blâme pour une action injuste ou non éthique sur autrui (Bartling et Fischbacher [2012]). En effet, on a tendance à être plus sévère avec celui qui prend la décision injuste qu'avec celui qui délègue cette décision. Le partage de responsabilités décourage également la rigueur morale (*e.g*., Falk et Szech [2013]). Le taux de mensonge augmente quand les actions sont complémentaires dans une équipe (Weisel et Shalvi [2015]) et avoir un complice potentiel, même sans possibilité de communiquer, augmente

le mensonge dans un jeu de dé simultané (Barr et Michailidou [2017]). De même, Falk et Szech [2017] ont conduit une expérience dans laquelle chaque membre d'un groupe de huit décide simultanément soit de recevoir de l'argent, soit de s'opposer au sacrifice de souris de laboratoire au prix du renoncement au gain. Il suffit qu'une personne fasse le choix égoïste pour que huit souris soient sacrifiées. Le résultat est que les choix égoïstes sont plus fréquents dans cette condition que dans la condition de base où les individus font leur choix de manière isolée. L'explication réside dans l'incertitude sur la responsabilité et la croyance (manipulable) selon laquelle l'individu est ou non pivot dans la détermination du résultat final.

Enfin, il est possible de se désengager moralement avec la proximité (« *bystander effect* », Darley et Latané [1968]). Cet effet explique que l'on n'interviendra pas pour aider une personne dans le besoin si d'autres personnes gravitent autour. En raison des interdépendances stratégiques, chacun laisse à autrui le soin d'agir et se sent moins coupable si personne n'intervient car la responsabilité est partagée (*e.g.*, Dana *et al*. [2007]). Le mécanisme permet plus de faiblesse morale que le mécanisme précédent puisqu'ici l'individu sait qu'en intervenant il règle le problème. Toutefois, Frommel *et al*. [2019] dans ce numéro montrent que la baisse de l'entraide à cause de cet effet ne heurte pas nécessairement la victime, par exemple si l'action d'une seule personne est suffisante, et ne résulte pas nécessairement d'un affaiblissement des normes sociales.

Récits narratifs et persuasion morale

Enfin, l'utilisation de *justifications* et de *récits* comme outils de persuasion morale regroupe un autre ensemble de stratégies qui visent à modifier les croyances en se fondant non pas sur ce que les gens font mais sur ce qu'ils disent. Les individus inventent des histoires pour se créer des excuses (Schiller [2017] ; Bénabou *et al*. [2018] ; Foerster et van der Weele [2018]). Dans Foerster et van der Weele [2018], un agent reçoit un signal bruité sur les externalités produites par une décision de don et entre en communication avec un autre agent ; les deux agents prennent ensuite une décision de don. Quand la décision est publique, un individu motivé par son image tend à minorer les externalités produites par son propre choix égoïste pour s'excuser (déni). Inversement, il majore les externalités en étant alarmiste pour convaincre l'autre agent de se comporter de manière pro-sociale.

Dans la même veine, Bénabou *et al.* [2018] montrent comment, pour garder une bonne image, les individus utilisent des récits dans leur raisonnement moral. Ils définissent le récit comme toute « nouvelle, histoire, expérience vécue ou heuristique qui a le pouvoir de modifier les croyances d'un agent sur le compromis entre les bénéfices privés et les coûts sociaux (ou l'inverse) auxquels fait face un individu, qui pourrait être lui-même, quelqu'un qu'il observe ou quelqu'un qu'il cherche à influencer » (p. 2). Ils opposent ces récits aux impératifs moraux qui sont des normes injonctives. Ces récits sont des signaux permettant de gérer les croyances sur les coûts et bénéfices moraux des actions. Les récits négatifs permettent de se créer des excuses pour mal se comporter tout en maintenant une bonne image, par exemple en sous-estimant les externalités de ses actions égoïstes, en augmentant artificiellement le coût supposé d'une conduite morale ou en considérant que son action individuelle n'est pas pivot. Les récits positifs, à l'opposé, maintiennent une pression sur l'individu pour renforcer sa responsabilité.

La dynamique morale de la société dépend des compromis entre ces types de récits. Un aspect essentiel de l'approche de Bénabou *et al* [2018] est que les récits se transmettent, y compris de manière virale, et peuvent ainsi générer une chaîne de comportements amoraux ou moraux, l'équilibre final dépendant de la composition du réseau. Une conclusion importante est que la mixité des interactions favorise les comportements pro-sociaux parce que les croyances deviennent moins polarisées.

CONCLUSION

Les recherches en économie comportementale et expérimentale de la dernière décennie ont révisé largement nos connaissances sur les mécanismes économiques et psychologiques qui gouvernent le raisonnement moral des individus, en particulier dans le domaine de l'honnêteté. Ces travaux ont montré que la plupart des individus résistent à la tentation de s'enrichir malhonnêtement, au moins partiellement, même quand il n'existe aucun risque de sanction, principalement parce qu'ils éprouvent une aversion intrinsèque au mensonge et parce que mentir peut porter atteinte à la façon dont ils sont perçus par autrui. Ces valeurs fondamentales permettent aux sociétés de se développer en relative harmonie.

Du point de vue théorique, le cadre classique de l'économie du crime beckerienne est compatible avec la prise en compte des coûts moraux du mensonge à côté des coûts monétaires de la tricherie. En revanche, il est nécessaire de sortir de ce cadre pour rendre compte des croyances de premier et de second ordre dans les choix éthiques. Or, les multiples expériences de laboratoire et de terrain récentes ont montré combien ces croyances sont importantes. D'autres cadres théoriques, s'appuyant notamment sur la théorie des jeux psychologiques, doivent donc être développés. Les modèles théoriques récents de Dufwenberg et Dufwenberg [2018], Gneezy *et al.* [2018], Abeler *et al*. [2019] et Khalmetski et Sliwka [2019] constituent une avancée importante sur cette voie. Ils sont complémentaires de la modélisation des préférences chez Bénabou et Tirole [2002] où l'utilité dépend d'une motivation monétaire, d'une motivation intrinsèque et d'une motivation d'image de soi (voir aussi Bénabou et Tirole [2006], et Bénabou et Tirole [2011]).

Un autre enseignement important des recherches en économie comportementale et expérimentale de la dernière décennie est une meilleure compréhension des stratégies mises en place par les individus pour gérer les dilemmes moraux de façon à s'enrichir sans attenter à leur image personnelle ni à leur image sociale, ou au contraire pour se créer davantage de contraintes s'ils craignent un manque de volonté. Ici également, la manipulation des croyances, la recherche d'incertitude, l'invention de récits sont fondamentales. Les travaux ont ainsi mis en évidence qu'au-delà des croyances motivées, maintes limitations cognitives des individus sont en fait également motivées. Rester ignorant des externalités négatives produites par ses actes non-éthiques, qu'ils relèvent d'un manque de générosité face à la détresse d'autrui ou d'un enrichissement malhonnête, permet de ne pas en subir le coût moral. Au contraire, mettre en évidence ces externalités négatives ou valoriser les externalités positives des actions guide l'individu sur un chemin vertueux.

Ces résultats ont des implications importantes non seulement pour la compréhension des comportements (non-)éthiques, mais également en termes de

politique économique de lutte contre la fraude et les incivilités. Ils ouvrent à une réflexion politique plus systématique sur l'action sur les croyances, l'incertitude, le dépistage des fausses informations. Ceci est d'autant plus crucial que les instruments classiques issus de la vision beckerienne de la fraude et de la tricherie - l'action sur la probabilité de détection et la sévérité des sanctions – n'ont pas toujours l'efficacité attendue (*e.g.*, Chalfin et McCrary [2017]) et ont même parfois des effets de diffusion inter-contextes négatifs à travers une pression négative sur l'honnêteté intrinsèque (Galeotti *et al*. [2019]).

Ces recherches en économie comportementale et expérimentale ouvrent la voie à de nouvelles investigations dans au moins deux directions. D'un côté, elles débouchent sur une réflexion totalement renouvelée sur les liens entre moralité et institutions, formelles (les règles et lois, les systèmes d'incitation, les modes d'organisation de la production et des marchés) et informelles (les normes injonctives et descriptives). D'un autre côté, elles appellent à de nouvelles recherches sur l'apprentissage de la moralité et des normes par des individus à la rationalité complète ou limitée.

RÉFÉRENCES BIBLIOGRAPHIQUES

ABELER J., BECKER A. et FALK A. [2014], « Representative Evidence on Lying Costs », *Journal of Public Economics*, 113, p. 96-104.

ABELER J., NOSENZO D. et RAYMOND C. [2019], « Preferences for Truth-Telling », *Econometrica,* 87(4), p. 1115-1153.

ALGER I. et WEIBULL J. [2013], « Homo Moralis.Preference Evolution under Incomplete Information and Assortative Matching », *Econometrica*, 8(6), p. 2269-2302.

ATTANASI G., BOUN MY K., GEORGANTZIS N. et GINES M. [2019], « Strategic ethics: Altruism without the other-regarding confound », *Revue Économique,* ce numéro.

ATTANASI G., RIMBAUD C. et VILLEVAL M.C. [2019], « Embezzlement and guilt aversion », *Journal of Economic Behavior & Organization,* 137, sous presse.

AZAR O.H., YOSEF S. et BAR ELI M. [2013], « Do customers return excessive change in a restaurant? A field experiment on dishonesty », *Journal of Economic Behavior & Organization*, 93, p. 219-226.

BALAFOUTAS L., BECK A, KERSCHBAMER R. et SUTTER M. [2013], « What drives taxi drivers? A field experiment on fraud in a market for credence goods », *Review of Economic Studies*, 80, p. 876-891.

BALASUBRAMANIAN P., BENNETT V.M. et PIERCE L. [2017], « The wages of dishonesty: The supply of cheating under high-powered incentives », *Journal of Economic Behavior & Organization,* 137, p. 428-444.

BANERJEE R., DATTA GUPTA N. et VILLEVAL M. C. [2018], « The Spillover Effects of Affirmative Action on Competitiveness and Unethical Behavior », *European Economic Review*, 101, p. 567-604.

BARR A. et MICHAILIDOU G. [2017], « Complicity without connection or communication », *Journal of Economic Behavior & Organization,* 142, p. 1-10.

BARTLING B. et FISCHBACHER U. [2012], « Shifting the blame: on delegation and responsibility », *Review of Economic Studies*, 79, p. 67-87.

BARTLING B., ENGL F. et WEBER R.A. [2014], « Does willful ignorance deflect punishment? An experimental study », *European Economic Review*, 70, p. 512-524.

BARTLING B., WEBER R.A. et YAO L. [2015], « Do markets erode social responsibility? », *Quarterly Journal of Economics*, 130, p. 219-266.

BARTLING B. et OZDEMIR Y. [2017], « The Limits to Moral Erosion in Markets: Social Norms and the Replacement Excuse », Working Paper University of Zurich.

BATSON C.D., THOMPSON E.R., SEUFERLING G., WHITNEY H. et STRONGMAN J.A. [1999], « Moral Hypocrisy: Appearing Moral to Oneself Without Being So », *Journal of Personality and Social Psychology*, 77(3), p. 525-537.

BATTIGALLI P. et DUFWENBERG M. [2007], « Guilt in games », *American Economic Review*, 97 (2), p. 170-176.

BAZERMAN M.H., TENBRUNSEL A.E. et WADE-BENZONI K. [1998], « Negotiating with yourself and losing: making decisions with competing internal preferences », *Academic Management Review*, 23(2), p. 225-241.

BECKER G.S. [1968], « Crime and punishment: An economic approach », *Journal of Political Economy,* 76, p. 169-217.

BELOT M. et SCHRÖDER M. [2013], « Sloppy work, lies and theft: A novel experimental design to study counterproductive behavior », *Journal of Economic Behavior & Organization,* 93, p. 233-238.

BÉNABOU R. [2015], « The Economics of Motivated Beliefs », *Revue d'Économie Politique*, 125(5), p. 665-685.

BÉNABOU R. et TIROLE J. [2002], « Self-confidence and personal motivation », *The Quarterly Journal of Economics*, 117(3), p. 871-915.

BÉNABOU R. et TIROLE J. [2006], « Incentives and Prosocial Behavior », *American Economic Review*, 96(5), p. 1652-1678.

BÉNABOU R. et TIROLE J. [2011], « Identity, Morals and Taboos: Beliefs as Assets », *The Quarterly Journal of Economics*, 126(2), p. 805-855.

BÉNABOU R., FALK A. et TIROLE J. [2018], « Narratives, Imperatives, and Moral Reasoning », CEPR Discussion Paper 13056.

BÉNISTANT J. et VILLEVAL M.C. [2019], « Unethical behavior and group identity in contests », *Journal of Economic Psychology,* sous presse.

BICCHIERI C. [2006], « *The grammar of society: The nature and dynamics of social norms* », Cambridge : Cambridge University Press.

BICCHIERI C. et XIAO E. [2009], «Do the right thing: But only if others do so », *Journal of Behavioral Decision Making*, 22(2), p. 191-208.

BICCHIERI C., DIMANT E., GÄCHTER S. et NOSENZO D. [2019a], « Observability, social proximity, and the erosion of norm compliance », Working Paper, SSRN https://ssrn.com/abstract=3355028.

BICCHIERI C., DIMANT E. et SONDEREGGER S. [2019b], « It's Not A Lie If You Believe It: On Norms, Lying, and Self-Serving Belief Distortion », Working Paper, SSRN https://ssrn.com/abstract=3326146

BICCHIERI C., DIMANT E. [2019], « Nudging with care: the risks and benefits of social information », *Public Choice*, sous presse.

BLANKEN I., VAN DE VEN N. et ZEELENBERG M. [2015], « A Meta-Analytic Review of Moral Licensing », *Personality and Social Psychology Bulletin*, 41(4), p. 1-19.

BØ. E., SLEMROD J. et THORESEN T. [2015], « Taxes on the Internet: Deterrence Effects of Public Disclosure », *American Economic Journal: Economic Policy*, 7(1), p. 36-62.

BUCCIOL A. et PIOVESAN M. [2011], « Luck or Cheating? A Field Experiment on Honesty With Children », *Journal of Economic Psychology*, 32 (1), p. 73-78.

BUSER T. et DREBER A. [2015], « The flipside of comparative payment schemes », *Management Science*, 62(9), p. 2626-2638.

CADSBY C.B., SONG F. et TAPON F. [2011], « Are you paying your employees to cheat ? An experimental investigation », *B.E. Journal of Economic Analysis & Policy*, 10(1), p. 1-32.

CADSBY C.B., DU N. et SONG F. [2016], « In-group favoritism and moral decision-making », *Journal of Economic Behavior & Organization,* 128, p. 59-71.

CAMERON J.S. et MILLER D.T. [2009], « Ethical Standards in Gain versus Loss Frames », In David de Cremer (Ed.), *Psychological Perspectives on Ethical Behavior and Decision Making*, Charlotte, Information Age Publishing, p.91-106.

CARLSON R.W., MARÉCHAL M.A., OUD B., FEHR E. et CROCKETT M. [2018], « Motivated misremembering: Selfish decisions are more generous in hindsight », Mimeo.

CARRELL S.E., MALMSTROM F.V. et WEST J.E. [2008], « Peer effects in academic cheating », *Journal of Human Resources,* 43(1), p.173-207.

CARRILLO J.D. et MARIOTTI T. [2000], « Strategic ignorance as a self-disciplining device », *Review of Economic Studies*, 67(3), p. 529-544.

CETTOLIN E., RIEDL A. et TRAN G. [2017], « Giving in the face of risk », *Journal of Risk and Uncertainty*, 55(2), p. 95-118.

CHALFIN A. et MCCRARY J. [2017], « Criminal Deterrence: A Review of the Literature », *Journal of Economic Literature*, 55(1), p. 5-48.

CHARNESS G. et DUFWENBERG, M. [2006], « Promises and Partnership », *Econometrica*, 74(6), p. 1579-1601.

CHARNESS G., MASCLET D. et VILLEVAL M.C. [2014], « The dark side of competition for status », *Management Science*, 60, p. 38-55.

CHARNESS G., BLANCO-JIMENEZ C., EZQUERRA L. et RODRIGUEZ-LARA I. [2019], « Cheating, Incentives, and Money Manipulation », *Experimental Economics*, 22 (1), p. 155-177.

CHARROIN L., FORTIN B., VILLEVAL M.C. [2019], « Homophily, peer effects and dishonesty », Miméo.

CHEW S.H., HUANG W. et ZHAO X. [2018], « Motivated false memory », Mimeo, National University of Singapore.

COFFMAN L.C. et GOTTHARD REAL A. [2019], « Moral Perceptions of Advised Actions », *Management Science,* sous presse.

COHN A., FEHR E. et MARÉCHAL M.A. [2014], « Business Culture and Dishonesty in the Banking Industry », *Nature*, 516 (7529), p. 86-89.

COHN A., MARÉCHAL M.A. et NOLL T. [2015], « Bad Boys: How Criminal Identity Salience Affects Rule Violation », *Review of Economic Studies*, 82 (4), p. 1289-1308.

COHN A. et MARÉCHAL M.A. [2019], « Laboratory Measure of Cheating Predicts School Misconduct », *The Economic Journal*, 128, p. 2743-2754.

COHN A., MARÉCHAL M.A., TANNENBAUM, D. et ZÜND, C.L. [2019], « Civic honesty around the globe », *Science*, 365(6448), p. 70-73.

COJOC D.C. et STOIAN A. . [2014], « Dishonesty and charitable behavior », *Experimental Economics*, 17, p. 717-732.

CONRADS J., IRLENBUSCH B., RILKE R.M. et WALKOWITZ G. [2013], « Lying and Team Incentives », *Journal of Economic Psychology*, 34, p. 1-7.

CONRADS J., IRLENBUSCH B., RILKE R.M., SCHIELKE A. et WALKOWITZ G. [2014], « Honesty in Tournaments », *Economics Letters*, 123 (1), p. 90-93.

CORICELLI G., JOFFILY M., MONTMARQUETTE C. et VILLEVAL M.C. [2010], «: Cheating, emotions and rationality: an experiment on tax evasion », *Experimental Economics*, 13, p. 226-247.

CORICELLI G., RUSCONI E. et VILLEVAL M.C. [2014], « Tax Evasion and emotions in repeated interactions: an empirical test of re-integrative shaming theory », *Journal of Economic Psychology*, 40, p. 49-61.

CRAWFORD, V.P. et SOBEL J. [1982], « Strategic information transmission », *Econometrica*, 50, 1431-1451.

DAI Z., GALEOTTI F. et VILLEVAL M.C. [2018], « Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation », *Management Science*, 64 (3), p. 1081–1100.

DANA J., WEBER R.A. et KUANG J.X. [2007], « Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness », *Economic Theory*, 33, p. 67-80.

DANILOV A., BIEMANN T., KRING T. et SLIWKA D. [2013], «: The dark side of team incentives: experimental evidence on advice quality from financial service professionals », *Journal of Economic Behavior & Organization*, 93, p. 266-272.

DARLEY J.M. et LATANÉ B. [1968], « Bystander Intervention in Emergencies: Diffusion of Responsibility », *Journal of Personality and Social Psychology*, 8 (4), p. 377-383.

DUFWENBERG M. et DUFWENBERG M.A. [2018], « Lies in Disguise—A Theoretical Analysis of Cheating », *Journal of Economic Theory*, 175, p. 248-264.

DUFWENBERG M. et MARTINSSON P. [2019], « Sealed Envelope Submissions Foster Research Integrity », *Revue Économique*, ce numéro.

DUGGAN M. et LEVITT S. [2002], « Winning Isn't Everything: Corruption in Sumo Wrestling », *American Economic Review*, 92(5), p. 1594-1605.

DULLECK U. et KERSCHBAMER R. [2006], « On doctors, mechanics, and computer specialists: The economics of credence goods », *Journal of Economic Literature*, 44, p. 542.

DULLECK U., KERSCHBAMER R. et SUTTER M. [2011], « The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition », *American Economic Review*, 101, p. 530-559.

EGAN M., MATVOS G. et SERU A. [2019], « The Market for Financial Adviser Misconduct », *Journal of Political Economy*, 127.

EHRLICH I. [1996], « Crime, Punishment and the Market for Offenses », *Journal of Economic Perspectives*, 10 (1), p. 43-67.

ERAT S. et GNEEZY U. [2012], « White lies », *Management Science*, 58(4), p. 723-733.

EXLEY C.L. [2015], « Excusing selfishness in charitable giving: The role of risk », *Review of Economic Studies*, 83, p. 587-628.

EXLEY C.L. et KESSLER J.B. [2018], « Motivated cognitive limitations », Mimeo, Harvard Business School.

FALK A. et SZECH N. [2013], « Morals and markets », *Science*, 340, p. 707-711.

FALK A. et SZECH N. [2017], « Diffusion of Being Pivotal and Immoral Outcomes », Discussion Paper, University of Bonn.

FARAVELLI M., FRIESEN L. et GANGADHARAN L. [2015], « Selection, tournaments, and dishonesty », *Journal of Economic Behavior & Organization*, 110, p. 160-175.

FISCHBACHER U. et FÖLLMI-HEUSI F. [2013], « Lies in Disguise—An Experimental Study on Cheating », *Journal of the European Economic Association*, 11 (3), p. 525-547.

FOERSTER M. et VAN DER WEELE J. [2018], « Denial and Alarmism in Collective Action Problems », Tinbergen Working Paper, University of Amsterdam.

FORTIN B., LACROIX G. et VILLEVAL M.C. [2007], « Tax evasion and social interactions », *Journal of Public Economics,* 91(11), p. 2089-2112.

FREDDI E. [2019], « Do People Avoid Morally Relevant Information? Evidence from the Refugee Crisis », *Review of Economics and Statistics,* sous presse.

FOSGAARD T.R., HANSEN L.G. et PIOVESAN M. [2013], « Separating will from grace: An experiment on conformity and awareness in cheating », *Journal of Economic Behavior & Organization*, 93, p. 279-284.

FROMELL H., NOSENZO D., OWENS T. et TUFANO F. [2019], « Are Victims Truly Worse Off in the Presence of Bystanders? Revisiting the Bystander Effect », *Revue Économique*, ce numéro.

GÄCHTER S. et SCHULZ J.F. [2016], « Intrinsic Honesty and the Prevalence of Rule Violations Across Societies », *Nature*, 531, p. 496-499.

GALEOTTI F., MAGGIAN V. et VILLEVAL M.C. [2019], « When Fraud Deterrence Institutions Reduce Intrinsic Honesty », Miméo.

GARBARINO E., SLONIM R. et VILLEVAL M.C. [2018], « A Method to Estimate Mean Lying Rates and Their Full Distribution », *Journal of the Economic Science Association,* 4(2), p. 136-150.

GARBARINO E., SLONIM R. et VILLEVAL M.C. [2019], « Loss Aversion and Lying Behavior: Theory, Estimation and Empirical Evidence », *Journal of Economic Behavior & Organization*, 158, p. 379-393.

GARCIA T., MASSONI S. et VILLEVAL M.C. [2019], « Ambiguity and excuse-driven behavior in charitable giving », SSRN, 3283773.

GEANAKOPLOS J., PEARCE D. et STACCHETTI E. [1989], « "Psychological Games and Sequential Rationality », *Games and Economic Behavior*, 1, p. 60-79.

GILL D., PROWSE V. et VLASSOPOULOS M. [2013], « Cheating in the Workplace: An Experimental Study of the Impact of Bonuses and Productivity », *Journal of Economic Behavior & Organization*, 96, p. 120-134.

GINO F., AYAL S. et ARIELY D. [2009], « Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel », *Psychological Science,* 20(3), p. 393-398.

GINO F., SCHWEITZER M.E., MEAD N.L. et ARIELY D. [2011], « Unable to resist temptation: How self-control depletion promotes unethical behavior », *Organizational Behavior and Human Decision Processes*, 115(2), p. 191-203.

GINO F., NORTON M. et WEBER R.A. [2016], « Motivated Bayesians: Feeling Moral While Acting Egoistically », *Journal of Economic Perspectives*, 30(3), p. 189-212.

GNEEZY U. [2005], « Deception: The Role of Consequences », *American Economic Review,* 95 (1), p. 384–394.

GNEEZY U., KAJACKAITE A. et SOBEL J. [2018], « Lying Aversion and the Size of the Lie », *American Economic Review*, 108 (2), p.419-453.

GNEEZY U., IMAS A. et MADARASZ K. [2014], « Conscience accounting: emotion dynamics and social behavior », *Management Science*, 60, p.2645-2658.

GRAVERT C. [2013], « How luck and performance affect stealing? », *Journal of Economic Behavior & Organization,* 93, p. 301-304.

GROLLEAU G., KOCHER M.G. et SUTAN A. [2016], « Cheating and loss aversion: do people lie more to avoid a loss? », *Management Science*, 62(12).

GROSS J., LEIB M., OFFERMAN T. et SHALVI S. [2018], « Ethical free-riding: When honest people find dishonest partners », *Psychological Science*, 29(12), p. 1956-1968.

GROSSMAN Z. [2014], « Strategic Ignorance and the Robustness of Social Preferences », *Management Science*, 60, p. 2659-2665.

GROSSMAN Z. et VAN DER WEELE J.J. [2017], « Self-image and willful ignorance in social decisions », *Journal of the European Economic Association*, 15(1), p. 173-217.

HAMMAN J., LOEWENSTEIN G. et WEBER R. [2010], « Self-interest through Delegation: An Additional Rationale for the Principal-Agent Relationship », *American Economic Review*, 100(4), p. 1826-1846.

HAISLEY E.C. et WEBER R.A. [2010], « Self-serving interpretations of ambiguity in other-regarding behavior », *Games and Economic Behavior*, 68, 614-625.

HOUSER D., VETTER S. et WINTER J. [2012], « Fairness and cheating », *European Economic Review*, 56(8), p. 1645-1655.

HOUSER D., LIST J.A., PIOVESAN M., SAMEK A. et WINTER J. [2016], « Dishonesty: From parents to children », *European Economic Review*, 82, p. 242-254.

IRLENBUSCH B. et VILLEVAL M.C. [2015], « Behavioral ethics: how psychology influenced economics and how economics might inform psychology? », *Current Opinions in Psychology,* 6, p. 87-92.

JACQUEMET N., LUCHINI S., ROSAZ J. et SHOGREN J.F. [2019], « Truth Telling Under Oath », *Management Science*, 65(1), Sous presse.

JACOB B.A. et LEVITT S. [2003], « Rotten apples: an investigation of the prevalence and predictors of teacher cheating », *The Quarterly Journal of Economics,* p. 843-877.

JACOBSEN C., FOSGAARD T.R. et PASCUAL-EZAMA D. [2017], « Why do we lie? A practical guide to the dishonesty literature », *Journal of Economic Surveys*, 32(2), p. 357-387.

JIANG T. [2013], « Cheating in Mind Games: The Subtlety of Rules Matters », *Journal of Economic Behavior & Organization*, 93, p. 328-336.

KAJACKAITE A. [2015], « If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment », *Journal of Economic Behavior & Organization*, 116, p. 518-524.

KAJACKAITE A. [2018], « Lying about luck versus lying about performance », *Journal of Economic Behavior & Organization*, 153, p. 194-199.

KAJACKAITE A. et GNEEZY U. [2017], « Incentives and Cheating », *Games and Economic Behavior*, 102, p. 433-444.

KAHNEMAN D. et TVERSKY A. [1979], « Prospect Theory: An Analysis of Decision Under Risk », *Econometrica*, 47, p. 263-291.

KARTIK N. [2009], « Strategic Communication With Lying Costs », *Review of Economic Studies*, 76 (4), p. 1359-1395.

KARTIK N., OTTAVIANI M. et SQUINTANI F. [2007], « Credulity, Lies, and Costly Talk », *Journal of Economic Theory*, 134 (1), p. 93-116.

KARTIK N., TERCIEUX O. et HOLDEN R. [2014], « Simple Mechanisms and Preferences for Honesty », *Games and Economic Behavior*, 83, p. 284-290.

KHALMETSKI K. et SLIWKA D. [2019], « Disguising Lies—Image Concerns and Partial Lying in Cheating Games », *American Economic Journal: Microeconomics*, sous presse.

KRÖLL M. et RUSTAGI D. [2016], « Shades of Dishonesty and Cheating in Informal Milk Markets in India », SAFE Working Paper, 134, Goethe University, Frankfurt.

KOUCHAKI M. et SMITH I.H. [2014], « The morning morality effect: The influence of time of day on unethical behavior », *Psychological Science*, 25(1), p. 95-102.

KOUCHAKI M. et GINO F. [2016], « Memories of unethical actions become obfuscated over time », *Proceedings of the National Academy of Sciences*, 113(22), p. 6166-6171.

LEE D.S. et MCCRARY J. [2016], « The Deterrence Effect of Prison: Dynamic Theory and Evidence », in *Advances in Econometrics*, *Regression Discontinuity Designs*, 38, Emerald Publishing, p. 73-146.

LEFEBVRE M., PESTIEAU P., RIEDL A et VILLEVAL M.C. [2015], « Tax evasion and social information: an experiment in Belgium, France, and the Netherlands. », *International Tax and Public Finance,* 22(3), p. 401-425.

LEVITT S.D. et LIST J.A. [2007], « What do laboratory experiments measuring social preferences tell us about the real world? », *Journal of Economic* Perspectives, 21(2), p. 153-174.

LIND J.T., NYBORG K. et PAULS A. [2016], « Save the planet or close your eyes? Testing strategic ignorance in a charity context », Miméo.

LÓPEZ-PÉREZ et SPIEGELMAN E. [2013], « Why Do People Tell the Truth? Experimental Evidence for Pure Lie Aversion », *Experimental Economics*, 16(3), p. 233-247.

LUNDQUIST T., ELLINGSEN T., GRIBBE E. et JOHANNESSON M. [2009], « The aversion to lying », *Journal of Economic Behavior &* Organization, 70, p. 81-92.

MAGGIAN V. et VILLEVAL M.C. [2006], « Social preferences and lying aversion in children », *Experimental Economics*, 19(3), p. 663-685.

MARÉCHAL M.A., COHN A., UGAZIO G. et RUFF C.C. [2017], « Increasing honesty in humans with noninvasive brain stimulation », *Proceedings of the National Academy of Sciences*, 114(17), p. 4360-4364.

MARTINELLI C., PARKER S.W., PÉREZ-GEA A.C. et RODRIGO R. [2018], « Cheating and Incentives: Learning from a Policy Experiment », *American Economic Journal: Economic Policy*, 10, p. 298-325.

MAZAR N., AMIR, O. et ARIELY D. [2008], « The Dishonesty of Honest People: A Theory of Self-Concept Maintenance », *Journal of Marketing Research*, 45(6), p. 633-644.

MEAD N.L., BAUMEISTER R.F., GINO F., SCHWEITZER M.E. et ARIELY D. [2009], « Too tired to tell the truth: Self-control resource depletion and dishonesty », *Journal of Experimental Social Psychology*, 45(3), p. 594-597.

MERRITT A.C., EFFRON D.A. et MONIN B. [2010], « Moral Self-Licensing: When Being Good Frees Us to Be Bad », *Social and Personality Psychology Compass*, 4(5), p. 344-357.

MONIN B. et JORDAN A.H. [2009], « The Dynamic Moral Self: A Social Psychological Perspective », In Narvaez D. et Lapsley D. (Eds.), *Personality, Identity, and Character: Explorations in Moral Psychology*. New York: Cambridge University Press.

MOORE C. [2016], « Always the hero to ourselves: the role of self-deception in unethical behavior », In: van Prooijen, J.W., van Lange, P. (Eds.), *Cheating, Corruption, and Concealment: the Roots of Dishonesty*. Cambridge: Cambridge University Press.

MULLAINATHAN S. [2002], « A memory-based model of bounded rationality », *The Quarterly Journal of Economics*, 117(3), p. 735-774.

NISAN M. [1991], « The moral balance model: Theory and research extending our understanding of moral choice and deviation », in KURTINE, W.M. et GEWIRTZ J.L. (Eds.), *Handbook of Moral Behavior and Development,* Lawrence Erlbaum Associates, Inc., p. 3-213.

PITTARELLO A., LEIB M., GORDON-HECKER T. et SHALVI S. [2015], « See what you want to see: justifications shape ethical blind spots », *Psychological Science*.

PLONER M. et REGNER T. [2013], « Self-Image and Moral Balancing: An Experimental Analysis », *Journal of Economic Behavior & Organization*, 93, p. 374-383.

POLIZZI DI SORRENTINO E., HERRMANN B. et VILLEVAL M.C. [2019], « Sweet truths and hungry lies? How metabolism affects (dis)honesty », Miméo.

POTTERS J. et STOOP J. [2016], « Do Cheaters in the Lab Also Cheat in the Field? », *European Economic Review*, 87, p. 26-33.

PRESTON I. et SZYMANSKI S. [2003], « Cheating in contests », *Oxford Review of Economic Policy*, 19(4), p. 612-624.

PRUCKNER G.J. et SAUSGRUBER R. [2013], « Honesty on the streets: A field study on newspaper purchasing », *Journal of the European Economic Association*, 11(3), p. 661-679.

QU E., MÉTÉREAU E., BUTERA L., VILLEVAL M.C. et DREHER J.C. [2019], « Neurocomputational mechanisms at play when weighing concerns for extrinsic rewards, moral values, and social image », *PLoS Biology,* 17(6), e3000283.

RAUHUT H. [2013], « Beliefs about Lying and Spreading of Dishonesty: Undetected Lies and Their Constructive and Destructive Social Dynamics in Dice Experiments», *PLoS One*, 8.11.

RUFFLE B.J. et TOBOL Y. [2014], « Honest on Mondays : Honesty and the temporal separation between decisions and payoffs », *European Economic Review*, 65, p. 126-135.

RUSTICHINI A. et VILLEVAL M.C. [2014], « Moral hypocrisy, power and social preferences », *Journal of Economic Behavior and Organization*, 107, p. 10-24.

SANDEL M.J. [2012], « *What money can't buy: The moral limits of markets*», Farrar, Straus et Giroux.

SAUCET, C., VILLEVAL, M.C. [2019], « Motivated Memory in Dictator Games », A paraître in *Games and Economic Behavior*.

SCHINDLER S. et PFATTHEICHER S. [2017], « The Frame of the Game: Loss-Framing Increases Dishonest Behavior », *Journal of Experimental Social Psychology*, 69, p. 172-177.

SHALVI S., DANA J., HANDGRAAF M. et DE DREU C. [2011], « Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptions and Behavior », *Organizational Behavior and Human Decision Processes*, 115 (2), p. 181-190.

SHALVI S., ELDAR O. et BEREBY-MEYER Y. [2012], « Honesty Requires Time (and Lack of Justifications) », *Psychological Science*, 23(10), p. 1264-1270.

SHALVI S. et DE DREU C. [2014], « Oxytocin promotes group-serving dishonesty », *Proceedings of the National Academy of Sciences*, 111, p. 5503-5507.

SHALVI S., GINO F., BARKAN R. et AYAL S. [2015], « Self-Serving Justifications: Doing Wrong and Feeling Moral », *Current Directions in Psychological Science*, 24(2), p. 125-130.

SHALVI S., SORRAPERRA I., VAN DER WEELE, VILLEVAL M.C. [2019], « Shooting the messenger: Supply and demand in the market for willful ignorance», Miméo.

SHILLER R. [2017], « Narrative Economics », *American Economic Review*, 107, p. 967-1004.

SHU L.L., GINO F. et BAZERMAN M.H. [2011], « Dishonest deed, clear conscience: when cheating leads to moral disengagement and motivated forgetting », *Personality and Social Psychology Bulletin*, 37(3), 330–349.

SHU L.L., MAZAR N., GINO F., ARIELY D. et BAZERMAN M.H. [2012], « Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end », *Proceedings of the National Academy of Sciences*, 109(38), p. 15197-15200.

TERGIMAN C. et VILLEVAL M.C. [2019], « The Way People Lie in Markets: the Role of Reputation and Competition », Miméo.

VANBERG C. [2008], « Why Do People Keep Their Promises? An Experimental Test of Two Explanations », *Econometrica*, 76(6), p. 1467-1480.

VAN DE VEN J. et VILLEVAL M.C. [2015], « Dishonesty under scrutiny », *Journal of the Economic Science Association,* 1, p. 86-99.

WANG J.T., SPEZIO M. et CAMERER C.F. [2010], « Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games », *American Economic Review*, 100(3), p. 984-1007.

WEISEL O. et SHALVI S. [2015], « The collaborative roots of corruption », *Proceedings of the National Academy of Sciences*, 112, 10651-10656.

WIBRAL M., DOHMEN T., KLINGMÜLLER D., WEBER B. et FALK A. [2012], « Testosterone Administration Reduces Lying in Men », *PLoS One*, 7 (10).

ZIMMERMANN F. [2019], « The dynamics of motivated beliefs », *American Economic Review, sous presse*.