



**HAL**  
open science

## Essays on the Behavioral Economics of Dishonesty.

Sorravich Kingsuwankul

► **To cite this version:**

Sorravich Kingsuwankul. Essays on the Behavioral Economics of Dishonesty.. Economics and Finance. Université Lumière - Lyon II, 2022. English. NNT : 2022LYO20001 . tel-04042864

**HAL Id: tel-04042864**

**<https://theses.hal.science/tel-04042864>**

Submitted on 23 Mar 2023

**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.



N° d'ordre NNT : 2022LYO20001

# THÈSE de DOCTORAT DE L'UNIVERSITÉ LUMIÈRE LYON 2

**École Doctorale : ED 486**

**Sciences économiques et de gestion**

Discipline : Sciences économiques

Soutenue publiquement le 7 septembre 2022, par :

**Sorravich KINGSUWANKUL**

---

## **Essays on the Behavioral Economics of Dishonesty.**

---

Devant le jury composé de :

Brice CORGNET, Professeur, EM Lyon Business School, Président

Nicolas JACQUEMET, Professeur des universités, Université Paris 1, Rapporteur

Lata GANGADHARAN, Professeure, Monash University, Rapporteur

Vanessa VALERO, Maîtresse de conférences, Loughborough University, Examinatrice

Marie-Claire VILLEVAL, Directrice de recherche, CNRS, Directrice de thèse

# Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « [Paternité – pas d'utilisation commerciale - pas de modification](#) » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.



# THÈSE DE DOCTORAT EN SCIENCES ÉCONOMIQUES

Opérée au sein de  
**L'Université Lumière Lyon 2**

École Doctorale no 486  
Sciences Économiques et de Gestion

**Spécialité de Doctorat** : Sciences Économiques

Soutenue publiquement le 7 Septembre 2022, par

**SORRAVICH KINGSUWANKUL**

---

## **Essays on the Behavioral Economics of Dishonesty**

Essais en économie comportementale de la malhonnêteté

---

Devant le jury composé de :

Brice CORGNET - Professeur, EM Lyon Business School (Examinateur)

Lata GANGADHARAN - Professeure, Monash University, Australie (Rapportrice)

Nicolas JACQUEMET - Professeur, Paris School of Economics (Rapporteur)

Vanessa VALERO - Lecturer, Loughborough University, Royaume-Uni (Examinatrice)

Marie Claire VILLEVAL - Directrice de recherche, CNRS (Directrice de thèse)



# THÈSE DE DOCTORAT EN SCIENCES ÉCONOMIQUES

Présentée en vue de l'obtention du grade de docteur de l'Université de Lyon,  
délivré par l'Université Lumière Lyon 2

---

## Essays on the Behavioral Economics of Dishonesty

Essais en économie comportementale de la malhonnêteté

---

Soutenue publiquement le 7 Septembre 2022, par

**SORRAVICH KINGSUWANKUL**

Directrice de thèse :

Marie Claire VILLEVAL - Directrice de recherche, CNRS, Université de Lyon

Jury :

Brice CORGNET - Professeur, EM Lyon Business School (Examinateur)

Lata GANGADHARAN - Professeure, Monash University, Australie (Rapporteuse)

Nicolas JACQUEMET - Professeur, Paris School of Economics (Rapporteur)

Vanessa VALERO - Lecturer, Loughborough University, Royaume-Uni (Examinatrice)

Université de Lyon - École Doctorale Sciences Économiques et de Gestion

Université Lumière Lyon 2

Groupe d'Analyse et de Théorie Économique, Lyon–St.Étienne



University Lumière Lyon 2 does not give any approbation or disapprobation about the thoughts expressed in this dissertation. They are only the author's ones and need to be considered as such.





# Acknowledgements

Spending nearly four years doing a PhD has been a challenging experience, scientifically and personally speaking. The journey was full of ups and downs, good and bad moments. I admit that there were many moments of doubts and discouragements, yet I chose to persevere. Along the way, I have learned many new things about research, the world and myself. It has been a rewarding experience not only because of the person who I have grown into, but also the people I have met along this memorable journey.

I am forever grateful to my supervisor, Marie Claire Villeval. I cannot fully describe in words my appreciation towards you, Marie Claire. You were the reason why I wanted to do a PhD in France. That dream came true. I thank you for all the attention you have constantly given me throughout my training. No matter how busy you are, you always have the time to supervise me, talk about research and the problems that I may have. It has been many long hours of scientific discussion and pep talks, and they have always been very fruitful and encouraging. Besides the great researcher that you are, I am extremely grateful for your kindness, generosity, and affection, which really comforted me during the difficult periods of doing the thesis. Lastly, thank you for having a strong faith in me and encouraged me to push my limits. I feel tremendously lucky to be your student and I truly admire you with great respect.

I thank my co-author, Chloe Tergiman, for her advice, expertise, and promptness. Chloe, it has been a pleasure and a great experience working with you. I really appreciate the opportunity and I have learned many things from you.

I thank Brice Corgnet, Lata Gangadharan, Nicolas Jacquemet and Vanessa Valero for having accepted to be the jury members of the thesis defense.

I extend my thanks to the senior members of GATE and GATE-Lab who were involved in the scientific activities during my training. Also, many thanks the administrative staff and the IT department who make sure everything at the lab is in order and functions well.

A big and special thank you to Quentin Thevenet for having spent many tremen-

---

dously long hours at the lab to teach me coding and to fix all the mess and bugs, especially the ones during an experiment. Thank you for always being present every time I run a session to assure my morale. Quentin, I dare say that you saved really many days of my PhD!

I am very glad to have met and become friends with the PhD students and post-docs at GATE who have all made this experience memorable and less lonely, especially during the period of Covid pandemic. In particular, and alphabetically listed – Aurélie, Claire, Charlotte, Eugenio, Fortuna, Jona, Jocelyn, Julien, Maria, Maxime, Morgan, Rémi, Siwar, Valentin, Vincent and Yao. I hope we will carry with us all the good memories we shared, especially those pizza and drinks we have had together. Special thanks to Jocelyn, Rémi, Tai and Tiruo who really smoothed the first few months of my arrival in France. Despite my experience living abroad, the adjustment period is never easy, but I felt welcomed at GATE and in France.

I am also very grateful to Thanee Chaiwat and Warn Lekfuangfu, both of whom I have worked with before starting my PhD. Thank you for the constant encouragement and the occasional calls now and then, where we talked both about the scientific and personal sides of life. Thank you for your kindness and for sharing your life experience with me.

I express deep gratitude to my parents who have always encouraged me to pursue my own passion and to be the person I envision myself to be. Thank you for all the education you gave me, which have opened so many doors. I owe every bit of success to both of you. I also thank Aui and Jew, my two brothers, for their love and support and the occasional virtual programming assistance. I hold great appreciation towards my grandmothers who taught me many important life skills.

Last but not the least, thank you François for all your support and positive thoughts during this journey. Thank you for sharing with me not only the joyful moments but also the stressful and tearful ones. You have added colors to my life.

# Résumé de la thèse

*“Ressentir tant pour autrui et si peu pour soi-même, au point d’en restreindre son égoïsme et de se satisfaire de notre affection bienveillante, constitue la perfection de la nature humaine...”*

— Adam Smith, La théorie des sentiments moraux

La moralité est inscrite en chacun de nous. Nous avons une boussole morale qui guide nos actions et nous avertit lorsque nous nous égarons *moralement*. Cette boussole fonctionne-t-elle si bien qu’elle nous éloigne de *tout* comportement immoral ? La suivons-nous toujours, ou fermons-nous parfois les yeux, volontairement ou non ? Les réponses à ces questions ne sont peut-être pas aussi simples qu’il n’y paraît.

Notre boussole morale nous donne souvent la bonne direction. Nous sommes conscients des normes sociales, c’est-à-dire des types de comportement que la plupart des gens considèrent comme moralement appropriés ou inappropriés. Bien que nous ayons une croyance commune qui prescrit les règles sociales non écrites, il nous arrive *de les violer*. Par exemple, le mensonge est considéré comme une chose inappropriée mais les gens continuent de mentir. Est-ce parce que notre boussole morale considère que certains mensonges sont *moins* immoraux que d’autres ? Ou est-ce seulement *nous* qui ignorons les règles en essayant de justifier certains types de mensonges ?

On peut faire valoir que le degré d’immoralité dépend des motifs qui sous-tendent ces mensonges. Les pieux mensonges, destinés à réduire les frictions sociales, sont souvent perçus comme moins problématiques (et donc moins inacceptables) car l’on peut penser que leur bénéfices l’emporte sur leur coût. C’est le cas, par exemple, lorsque nous enjolivons la vérité auprès de nos proches (comme dire à notre partenaire que la couleur de sa tenue est assortie à son teint) où lorsque les parents disent aux enfants de bien se comporter avant Noël (sinon le Père Noël n’apportera pas de cadeaux). Au-delà de ces mensonges apparemment inoffensifs, il y a les mensonges intéressés qui apportent des avantages personnels au détriment d’autrui. Dans notre vie quotidienne, nous sommes confrontés à des occasions de nous enrichir par des moyens contraires à l’éthique - qu’il s’agisse de petites choses, comme lorsque nous trouvons un portefeuille perdu dans la rue, ou de choses plus importantes ayant

des conséquences plus graves, comme l'évasion fiscale et le détournement de fonds. C'est de tels écarts de conduite que notre boussole morale tente d'éviter. Pourtant, il nous arrive d'ignorer cette boussole, consciemment ou non. Nous essayons alors de justifier ces entorses à l'éthique pour rester en paix avec notre propre conscience.

Nous sommes le produit de notre propre évolution, et il en va de même pour notre perception de ce qui est bien ou mal. Au fur et à mesure que les sociétés progressent dans le temps, les individus apprennent les règles normatives. Les actions qui apportent un bienfait social sont récompensées, tandis que celles qui lui portent préjudice sont sanctionnées. Cela conduit à une construction de standards normatifs où les actes vertueux sont renforcés et perpétués, et où les actes immoraux sont réprimés et abandonnés. Les individus intériorisent ensuite ces règles sociales et, quand leurs enfants voient le jour, ces vertus sont transmises de génération en génération. Le processus se poursuit ainsi, façonnant l'ensemble des règles normatives de nos sociétés actuelles. En bref, le compas moral (notre sens de la moralité) est le produit d'une coévolution autant biologique que culturelle ([Gintis et al., 2005](#); [Gintis and Helbing, 2015](#)).

Bien que nous ayons intériorisé ces normes, nous pouvons parfois être tentés de les enfreindre et de nous comporter de manière égoïste. Faire cela crée en nous une dissonance cognitive causée par un décalage entre notre sens de la moralité et notre propre comportement. Nous pouvons nous sentir intérieurement perturbés par des émotions négatives telles que la culpabilité et la honte, car nous sommes conscients d'avoir dévié de ce qui est considéré comme un comportement moralement adapté. Ces *coûts psychiques* peuvent être importants lorsque notre méfait cause du tort aux autres. En outre, ils peuvent être amplifiés lorsque nous sommes exposés au jugement des autres, qui attendent de nous que nous respections les règles, et nous le savons. Par conséquent, lorsque les individus violent les règles normatives, qu'ils ont eux-mêmes construites et intériorisées, ils subissent des *coûts moraux* découlant de la violation de leurs motivations intrinsèques.

## Objectifs

Cette thèse se situe à l'intersection de la morale et de l'économie. Elle contribue au domaine en constante expansion de l'économie comportementale de la malhonnêteté en fournissant des preuves inédites de l'influence des coûts moraux sur le mensonge et les décisions économiques. S'appuyant sur la méthodologie de l'économie expérimentale, elle vise à tester la mesure dans laquelle les coûts moraux, découlant des motivations individuelles intrinsèques, peuvent limiter les comportements mensongers, dans quelles circonstances leur influence peut être limitée, et s'ils

peuvent affecter la façon dont les gens prennent des décisions économiques.

En bref, les chapitres 1 et 2 se concentrent sur la compréhension de l'influence du coût moral pour limiter les comportements mensongers dans les contextes respectifs des marchés financiers et de groupe. Le chapitre 3 adopte une autre perspective en s'interrogeant sur la persistance de l'effet du coût moral sur le processus de décision *après* avoir menti. Essentiellement, il vise à comprendre si le coût moral du mensonge influence la façon dont les gens utilisent leur argent gagné de manière non éthique.

Plus précisément,

- Le chapitre 1 étudie l'impact d'un serment d'honnêteté sur le comportement mensonger dans un contexte de marché financier à l'aide d'une expérience de laboratoire.
- Le chapitre 2 examine l'impact d'un serment individuel contre le mensonge dans un contexte de groupe en utilisant une expérience de laboratoire.
- Le chapitre 3 cherche à savoir si les gens traitent l'argent malhonnête gagné au prix d'un coût moral plutôt comme un gain inattendu que comme le résultat d'un effort coûteux dans le contexte de la prise de risque en utilisant une expérience en ligne.

La morale a sa place en économie. La théorie des sentiments moraux d'Adam Smith a fourni un fondement moral à *l'amélioration de la société humaine* : comment pouvons-nous, en tant que créatures sociales, prospérer et vivre en harmonie les uns avec les autres ? Les idées de Smith sont très proches des travaux sur les normes sociales (voir [Elster, 1989](#); [Ostrom, 2000](#); [Bicchieri, 2005](#)). Nous apprenons les normes des autres, et vice-versa. Les actions moralement appropriées sont renforcées par l'approbation sociale, tandis que celles jugées moralement inappropriées sont dissuadées par des sanctions sociales. Ces règles sont assimilées par notre conscience morale qui nous guide automatiquement, nous indiquant ce qui est bien ou mal. Nous ne cherchons pas seulement à obtenir le meilleur pour nous-mêmes, mais aussi à répondre aux attentes des autres, ces dernières étant régies par un ensemble de règles juridiques et sociales - écrites ou non. Toutefois, nous pouvons *céder* à la recherche d'une *auto promotion éhontée*, ce qui nous amène à trouver des justifications égoïstes pour valider nos actions douteuses avant de les entreprendre ou même après les avoir entreprises.

Parce que nous vivons dans une société avec des règles normatives, nous avons développé des *préférences sociales* et *morales*, en plus de nos désirs d'atteindre nos propres objectifs ([Gintis et al., 2005](#)). Ces idées montrent la perspective des

économistes comportementalistes sur le sujet des prises de décision humaines. Nous nous soucions du bien-être des autres - nous les traitons de manière équitable (ex. [Rabin, 1993](#); [Fehr and Schmidt, 1999](#); [Charness and Rabin, 2002](#); [Falk et al., 2003](#)), nous prêtons attention à la manière dont un résultat est atteint (ex. [Frey et al., 2004](#); [Falk and Fischbacher, 2006](#); [Falk et al., 2008](#)), répondons à la gentillesse par de la gentillesse, et punissons ceux qui transgressent les règles, à notre dépen ou au dépen des autres (ex. [Fehr and Gächter, 1998](#); [Abbink et al., 2000](#); [Fehr and Gächter, 2000](#); [Masclat et al., 2003](#); [Fehr and Fischbacher, 2004](#)). Outre notre préférence pour la considération d'autrui, nous sommes intrinsèquement motivés à traiter les gens avec décence et respect, sans aucune intention maligne, simplement parce que nous pensons que *c'est la bonne chose à faire* ([Gintis, 2016](#)). Ces motivations innées englobent des caractéristiques vertueuses telles que l'honnêteté, la loyauté, la fiabilité, la diligence et le courage. Elles sont poursuivies, du moins en partie, pour *leur intérêt propre* plutôt que pour *leurs effets*.

Néanmoins, nous pouvons être amenés à faire des compromis entre ces motivations intrinsèques et des motivations liées à notre propre intérêt ou à celui d'autrui, en fonction du coût de l'action (im)morale pour nous-mêmes et des avantages ou inconvénients pour autrui. Un tel compromis diminue la "valeur" de l'honnêteté. Ainsi, même si nous sommes guidés par des règles normatives, que nous nous soucions des autres membres de la communauté et que nous avons le désir inné d'être moraux, nous sommes parfois confrontés à des dilemmes moraux. Un point important, cependant, est *qu'au moins* nous essayons d'être *moraux*. Ainsi, la façon dont nous prenons des décisions dans un domaine moral nous rapproche davantage d'un *homo moralis* ([Gintis and Helbing, 2015](#)) – un homme qui a un désir inné d'agir moralement - qu'à un *homo economicus* – un homme qui fonde sa décision sur un simple calcul des coûts et des avantages.

Le cadre théorique de [Becker \(1968\)](#) applique une analyse coûts-bénéfices à un comportement criminel, ce qui suggère qu'une personne rationnelle va mentir (par exemple faire une fausse déclaration d'impôt sur le revenu, escroquer des clients. . .) si les gains financiers dudit mensonge sont supérieurs aux coûts associés au fait d'être pris. Bien qu'elle ne rejette pas le rôle des motivations intrinsèques en tant que déterminants des prises de décision non éthiques, elle offre une perspective plutôt froide des comportements (im)moraux. Cependant, les modèles économiques standard de la criminalité ne parviennent pas à expliquer certains schémas comportementaux de tricherie individuelle. Premièrement, même en l'absence de toute détection, il existe une hétérogénéité dans le comportement moral des individus. Notamment, le comportement de tricherie est généralement inférieur à ce que le modèle économique standard pourrait prédire. En fait, certains individus ne trichent tout

simplement pas. Cela signifie que les gens ne traitent pas les décisions contraires à l'éthique comme un pari et qu'ils hésitent à tricher, révélant ainsi l'existence de *coûts moraux* (voir [Abeler et al., 2014, 2019](#)). Deuxièmement, il existe une malhonnêteté incomplète, dans laquelle les gens *ne trichent pas* aussi amplement qu'ils le pourraient, mais le font plutôt *partiellement*. Cela suggère que les coûts moraux sont non linéaires et que les gens peuvent se soucier de la façon dont ils sont perçus par les autres (voir [Fischbacher and Föllmi-Heusi, 2013](#)). Ces deux modèles comportementaux ne peuvent être expliqués par le modèle économique standard, ce qui illustre la nécessité d'étendre la modélisation économique afin d'intégrer les préférences *morales* et *sociales* dans la fonction d'utilité individuelle. Cela constitue justement la contribution des comportementalistes. Parmi les travaux les plus importants, citons les modèles d'aversion pour la culpabilité (voir [Battigalli and Dufwenberg, 2007](#)), les normes sociales (voir [Krupka and Weber, 2013](#)) et l'aversion pour la tricherie perçue (voir [Gneezy et al., 2018](#); [Dufwenberg and Dufwenberg, 2018](#); [Khalmetski and Sliwka, 2019](#); [Abeler et al., 2019](#)). Selon le modèle d'[Abeler et al. \(2019\)](#), un individu veut être honnête et perçu comme tel, chaque fois que son action s'écarte de ces motivations, il encourt un coût moral du mensonge. Cette thèse contribue à apporter des preuves empiriques sur l'économie comportementale de la malhonnêteté en se concentrant sur la recherche d'un coût moral dans le processus de décision individuel.

En résumé, les données empiriques sur l'économie comportementale de la malhonnêteté suggèrent que tout le monde n'est pas un *homo economicus* lorsqu'il s'agit de prendre des décisions dans le domaine de la morale. Les gens ont des motivations intrinsèques qui induisent des coûts moraux lorsque leurs actions ne sont pas alignées sur leurs règles internes. En outre, les comportementalistes ont montré que le comportement mensonger est sensible à la manipulation exogène des coûts moraux tels que le serment d'honnêteté, qui fait l'objet des chapitres 1 et 2 de cette thèse.

L'un des domaines inexplorés est l'impact d'un serment d'honnêteté dans un contexte d'interactions sociales où le mensonge est stratégique et crée une externalité négative sur un autre individu. Bien que [Jacquemot et al. \(2019, 2021\)](#) aient exploré un tel cadre à l'aide d'un jeu de communication stratégique à information asymétrique (Sender-Receiver Game), il ne s'agissait que d'une décision unique et ponctuelle. Néanmoins, dans un contexte tel que celui d'un marché financier, les conseillers interagissent souvent avec les investisseurs de manière répétée et ont ainsi plusieurs occasions de les escroquer en exploitant l'asymétrie d'information. Cette interaction répétée, en plus de correspondre plus étroitement à la dynamique du monde réel, peut fournir des indications intéressantes sur l'effet d'un serment où



les investisseurs peuvent évaluer l'honnêteté de leurs conseillers tandis que les conseillers peuvent adapter leurs stratégies et leurs mensonges au fil du temps. En outre, la littérature existante n'a pas abordé l'impact d'un serment d'honnêteté sur la décision d'une contrepartie qui est victime de mensonge. Dans un contexte de marché financier, si les investisseurs savent que les conseillers sont sous serment, seraient-ils plus disposés à suivre leurs conseils ? Comment un serment d'honnêteté fonctionne-t-il dans le cadre d'un tel marché ? - Quelles sont les conséquences de l'introduction d'un serment sur les comportements mensongers des conseillers ? Comment cela affecte-t-il finalement le bien-être des investisseurs ? Le chapitre 1 comble cette lacune de la littérature.

### **Chapitre 1 : Les serments d'honnêteté dans les relations conseiller-client: une expérience**

L'objectif du chapitre 1 est d'identifier ce qui fait l'efficacité des serments d'honnêteté dans le cadre des marchés financiers en utilisant une expérience de laboratoire. On peut s'interroger sur la pertinence de la prestation de serment dans ce cadre particulier. *Le serment du banquier* a été proposé pour limiter les fautes professionnelles et restaurer la confiance du public dans le secteur bancaire, celle-ci ayant diminué à la suite de la crise financière mondiale de 2008 (Sapienza and Zingales, 2012). À l'instar du serment d'Hippocrate pour la profession médicale, le serment du banquier engage les employés du secteur financier à respecter les normes éthiques prescrites, à maintenir transparence et responsabilité, et à placer l'intérêt du client avant les intérêts personnels (Boatright, 2013). Cette pratique a été légalement mise en œuvre sur tous les employés du secteur bancaire aux Pays-Bas depuis 2015, suivis par la Belgique en 2019. Comme la littérature l'a montré, une telle pratique peut s'avérer être un outil rentable pour réduire les comportements mensongers.

Sur un marché financier, l'investisseur ne sait pas si le produit recommandé par le conseiller est effectivement dans son intérêt. En tant qu'expert disposant d'informations supérieures, le conseiller peut escroquer l'investisseur en lui présentant de manière inexacte la situation réelle du monde (c.a.d le portefeuille). De son côté, le conseiller peut envisager différentes stratégies de mensonge - certaines qui sont *décelables ex-post* par l'investisseur avec des probabilités variées, et d'autres qui restent *niables* (en d'autres termes, qui ne peuvent jamais être détectées). Du côté de l'investisseur, après avoir pris sa décision et observé le résultat apporté par l'investissement, il peut être en mesure de déduire l'honnêteté du conseiller. Il s'agit d'un aspect important car, si l'investisseur peut déduire qu'on lui a menti, cela peut avoir une incidence négative sur sa décision de participer au marché financier, ce qui conduit au problème du *marché des citrons* ("lemons market") (Akerlof, 1970).

Par conséquent, si un serment peut préserver l'honnêteté des conseillers, il peut contribuer à empêcher les investisseurs d'en sortir.

Plus précisément, le chapitre 1 étudie de manière expérimentale l'impact d'un serment d'honnêteté sur le comportement mensonger des conseillers financiers, et par conséquent son impact sur les investisseurs dans un cadre contrôlé en laboratoire. Le serment élimine-t-il *tous* les types de mensonges ou change-t-il simplement la façon dont les conseillers mentent à leurs investisseurs ? Les investisseurs font-ils davantage confiance aux conseillers en raison du serment *en tant que tel* ? Ou est-ce l'expérience passée avec leurs conseillers qui dicte les décisions d'investissement ?

La principale contribution du chapitre 1 est d'étudier l'impact du serment *au-delà* de la réduction des mensonges - comment cela influence la prise de décision de l'investisseur victime d'un mensonge. [Weitzel and Kirchler \(2021\)](#) ont mené une étude d'audit pour examiner l'impact du serment du banquier sur l'honnêteté des conseillers aux Pays Bas. Ils ont constaté que le fait de se voir rappeler le serment rendait les conseillers moins susceptibles de privilégier les intérêts de la banque (c'est-à-dire d'accorder des prêts inutiles) que dans le scénario de base, ce qui montre qu'un souvenir atténué du serment pourrait jouer un rôle. C'est aussi le cas pour les rappels moraux (les banquiers devaient expliquer l'objectif du serment) et pour la peur de la sanction (les clients et l'institution elle-même pouvaient être plus susceptibles de déposer des plaintes pour mauvaise conduite). Toutefois, la question de l'impact du serment sur les décisions des investisseurs reste ouverte : le fait de savoir que le conseiller est sous serment facilite-t-il la décision de l'investisseur ? Ou est-ce l'expérience passée avec le conseiller qui compte le plus pour lui ? Le chapitre 1 explore ces phénomènes par le biais d'un design expérimental inédit faisant varier le fait que le serment soit connu ou non des investisseurs.

En outre, la plupart des travaux existants se sont concentrés sur une tâche de mensonge individuel non stratégique et aucun n'a spécifiquement examiné l'impact du serment dans un cadre stratégique où les mensonges génèrent une externalité négative - deux aspects importants sur les marchés financiers. Bien que certaines études antérieures aient utilisé un jeu de type "Sender-Receiver" (ex. [Jacquemet et al., 2019, 2021](#)), celui-ci est mis en œuvre comme un jeu à coup unique où le récepteur (la victime du mensonge) n'apprend pas par la suite si le message était effectivement véridique ou non. En revanche, dans un contexte de marché financier, après avoir suivi les recommandations du conseiller, l'investisseur peut découvrir a posteriori qu'on lui a menti. Cela crée une dynamique d'apprentissage dans le cadre d'un jeu répété, par laquelle les investisseurs peuvent commencer à perdre confiance en leurs conseillers.

Nous avons conçu une expérience de laboratoire utilisant le jeu d'annonce développé par [Tergiman and Villeval \(2022\)](#). Dans ce jeu à deux joueurs, un conseiller observe en privé un portefeuille de projets d'investissement, représentés par trois cartes. Chaque carte a une probabilité de 0,5 d'afficher une étoile, signifiant un projet réussi. Dans une situation d'asymétrie de l'information, le conseiller envoie alors un message en conversation libre ("cheap talk") à un joueur jouant le rôle de l'investisseur à propos du nombre d'étoiles de son portefeuille. L'investisseur décide alors d'investir ou non dans le projet. Ensuite, la Nature tire une des trois cartes du conseiller pour déterminer le succès de l'investissement (une étoile ou aucune étoile). Dans ce contexte, le conseiller peut choisir de mentir ou non et de choisir quel type de mensonge utiliser. Un mensonge peut être détecté par l'investisseur (par exemple, annoncer 3 étoiles après en avoir observé moins de 3) ou nié (par exemple, annoncer 2 étoiles après en avoir observé moins de 2). Ainsi, alors que les mensonges détectables peuvent être découverts a posteriori, les mensonges niables ne peuvent pas être découverts *immédiatement* car le tirage de la Nature ne contredit pas l'annonce du conseiller. Nous avons utilisé un protocole dit de "Stranger Matching" où les conseillers et les investisseurs sont ré-appariés après chaque période.<sup>1</sup>

Nos quatre traitements inter-sujets varient en fonction de l'introduction ou non d'un serment et de la manière dont celui-ci est présenté. L'absence de serment (*No Oath*) sert de condition de base (comme décrit dans le paragraphe précédent). Dans le cas d'un serment connu de tous (*Oath-Common-Knowledge*), les conseillers prêtent serment d'honnêteté et cela est connu de tous (conseillers et investisseurs). La comparaison de Oath-CK avec la condition de base nous permet d'évaluer directement l'impact d'un serment connu par tous. Nous avons constaté une réduction drastique des mensonges des conseillers (détectables et niables), ce qui a entraîné une augmentation des investissements.

Cependant, un taux d'investissement plus élevé peut provenir de deux facteurs. D'une part, il est possible que les investisseurs anticipent que les annonces des conseillers soient plus honnêtes car ils connaissent l'existence du serment. En d'autres termes, la connaissance du serment a une valeur informative pour les investisseurs, leur permettant de faire davantage confiance aux annonces que dans le cas où ils ignorent l'existence dudit serment. D'autre part, étant donné que les annonces faites sous un serment connu de tous sont essentiellement honnêtes, il est possible que les investisseurs réagissent simplement au comportement plus honnête des conseillers durant l'expérience. Les deux autres traitements, *Oath-No-Common-Knowledge* et *Oath-No-Common Knowledge-Asynchronous*, permettent de distinguer les deux ex-

---

<sup>1</sup>Bien qu'il soit possible de détecter un menteur utilisant le mensonge niable *au fil du temps*, il est presque impossible de le faire sous le protocole de "Stranger Matching."

plications.

Dans Oath-NoCK, les conseillers prêtent serment mais les investisseurs ne le savent pas. Dans Oath-NoCK-Async, tous les participants sont des investisseurs n'ayant pas connaissance du serment. Ils sont confrontés aux annonces faites auparavant par des conseillers du traitement Oath-CK.

Les résultats confirment l'explication de l'expérience. Nous avons constaté que c'est au constat de *l'honnêteté de leurs conseillers* que les investisseurs réagissent et non au fait qu'ils savent que les conseillers sont sous serment. En d'autres termes, les investisseurs apprennent l'honnêteté de leurs conseillers grâce à leur expérience. Si la valeur informative du serment joue un rôle marginal dans la décision des investisseurs, elle est un élément crucial pour les conseillers. Nous avons constaté que les conseillers mentent davantage lorsque le serment n'est pas connu des investisseurs, ce qui entraîne un niveau d'honnêteté moindre qui ne permet pas de maintenir la confiance. En d'autres termes, il est important de rappeler aux conseillers que les investisseurs sont au courant du serment. Nos conclusions sont particulièrement importantes étant donné que notre enquête indépendante en ligne menée via Prolific auprès de 200 résidents des Pays-Bas, où le serment du banquier est légalement contraignant, indique que 68% d'entre eux ne savaient pas que les banquiers néerlandais étaient tenus de prêter serment. Ainsi, les conclusions du chapitre 1 ont des implications concernant la mise en œuvre du serment ou des codes de conduite professionnels : elle doit s'accompagner d'une stratégie de relations publiques cohérente et bien planifiée pour garantir une sensibilisation à l'intérieur comme à l'extérieur de l'organisation.

Dans l'ensemble, le chapitre 1 a exploré l'impact d'un serment d'honnêteté sur le comportement mensonger dans un cadre stratégique avec des interactions sociales et une externalité négative - un cadre s'approchant de celui des marchés financiers. Notre contribution inédite consiste à faire varier le fait qu'un serment soit connu ou non de tous (conseillers comme investisseurs) pour examiner son impact sur les comportements mensongers et par conséquent sur les décisions d'investissement. Nous avons identifié que même si le serment en lui-même a une faible valeur informative pour les investisseurs, le fait qu'il soit connu de tous est une condition nécessaire pour soutenir l'honnêteté des conseillers afin de créer une expérience qui maintienne la confiance des investisseurs.

Un autre domaine inexploré dans la littérature sur la prestation de serment est l'impact d'un serment d'honnêteté lorsque les individus se trouvent dans un contexte de groupe. Étant donné qu'un serment est adopté par les praticiens pour réduire les fautes professionnelles au sein d'une organisation, il est important d'examiner si une telle intervention est efficace dans *un contexte collectif* où les individus peu-

vent communiquer entre eux et peuvent avoir besoin de prendre conjointement des décisions éthiques. Le chapitre 2 est le premier à étudier ce phénomène dans le cadre d'une expérience de laboratoire.

## Chapitre 2 : Serment individuel et mensonge en groupe

Le chapitre 2 poursuit l'exploration de l'impact du serment d'honnêteté d'un autre point de vue. Il vise à vérifier si un serment réduit le mensonge lorsque les individus se trouvent dans un contexte de groupe. Il s'agit d'un contexte pertinent à étudier pour deux raisons. Premièrement, les fraudes et les scandales financiers impliquent souvent *des groupes d'individus*, et non *une seule personne*. Enrons, l'affaire Madoff ou Wells Fargo sont quelques exemples où des groupes d'individus sont impliqués dans des réseaux de crimes financiers complexes. En second lieu, après qu'une organisation ait mis en place un serment ou un code de conduite professionnel pour ses employés, ceux-ci se retrouveront dans un environnement de travail où ils pourront être confrontés à un objectif commun et amenés à prendre des décisions éthiques avec d'autres personnes au sein de leur organisation. Un serment d'honnêteté découragerait-il efficacement les comportements malhonnêtes dans un tel contexte de groupe ?

La littérature sur le comportement mensonger suggère que les groupes sont plus malhonnêtes que les individus (ex. [Weisel and Shalvi, 2015](#); [Kocher et al., 2018](#)). Plusieurs raisons expliquent ce *glissement vers la malhonnêteté* (“dishonesty shift”) dans les groupes. Tout d'abord, les études sur la prise de décision en groupe ont révélé que les groupes sont plus *sophistiqués* que les individus et sont plus susceptibles de faire un choix intéressé lors d'un jeu économique (voir la synthèse de [Charness and Sutter, 2012](#)). Par exemple, les groupes choisissent un nombre plus bas dans un jeu de concours de beauté ([Kocher and Sutter, 2005](#)), envoient et retournent moins d'argent dans un jeu de confiance (ex. [Cox, 2002](#); [Kugler et al., 2007](#)), font et acceptent une offre plus petite dans un jeu d'ultimatum ([Bornstein and Yaniv, 1998](#)). En ce qui concerne le mensonge, [Sutter \(2009\)](#) a constaté dans un jeu d'envoi et de réception (Sender-Receiver Game) que les groupes sont plus susceptibles d'utiliser une tromperie sophistiquée que les individus (le participant ayant le rôle du “Sender” dit la vérité tout en croyant que son indication ne sera pas suivi). Deuxièmement, le contexte du groupe permet la *diffusion de la responsabilité* (ex. [Falk and Szech, 2013](#); [Falk et al., 2020](#)). Ceci se produit lorsque l'action d'un individu envers une décision de groupe devient moins déterminante, ce qui entraîne une diminution de la perception de la responsabilité individuelle. Ainsi, les individus exploitent une marge de manœuvre morale ([Dana et al., 2007](#)) en faveur de leur action immorale en se cachant derrière les autres (ex. [Conrads et al., 2013](#); [Weisel and Shalvi, 2015](#);

Bauer et al., 2021). Troisièmement, les groupes peuvent mentir davantage que les individus parce que cela peut *bénéficier aux autres* (ex. Wiltermuth, 2011; Gino et al., 2013). Cependant, une étude récente de Kocher et al. (2018) a révélé que les groupes mentent dans la même mesure, indépendamment du fait que d'autres puissent en bénéficier ou non. Enfin, des études antérieures suggèrent que les individus d'un groupe ont tendance à mentir davantage parce que ce cadre leur permet d'apprendre la norme empirique de l'honnêteté (c'est à dire ce que font les autres) en observant l'action de leurs pairs (Bicchieri et al., 2022) ou en communiquant avec les autres (Kocher et al., 2018). Autrement dit, le virage malhonnête dans les groupes résulte de *l'érosion de la norme d'honnêteté*. Le chapitre 2 se concentre sur les deux dernières explications : comment fonctionnerait un serment lorsque les individus peuvent communiquer avec les autres membres de leur groupe, et comment cela interagit-il avec la structure incitative ?

La contribution du chapitre 2 est double. Étant donné que la littérature antérieure a montré qu'un serment d'honnêteté est efficace pour réduire le comportement mensonger au niveau individuel, ce chapitre est le premier à examiner si cet impact *persiste* lorsque les individus sont dans un groupe. D'une part, le serment augmente le coût moral du mensonge et rend la norme d'honnêteté plus saillante. Un individu peut donc promouvoir un comportement honnête au sein de son groupe. D'autre part, comme les membres du groupe peuvent communiquer, ils peuvent échanger des justifications et se convaincre mutuellement de rompre le serment. Cela est possible puisque la littérature sur la conformité des pairs a montré que les individus ont tendance à mentir par conformisme après avoir observé les violations de normes de leurs pairs (voir ex. Gino et al., 2009; Fosgaard et al., 2013; Innes and Mitra, 2013; Kroher and Wolbring, 2015). Ainsi, ce chapitre apporte une contribution tant au sujet de l'effet de conformité par les pairs que celui du mensonge sous serment dans un contexte de groupe. En outre, ce chapitre cherche à savoir si une intervention sous serment interagit avec la structure incitative à laquelle sont confrontés les membres du groupe. Cette question est pertinente pour les organisations qui imposent un serment à leurs employés, car les incitations du groupe peuvent interagir avec cette intervention de telle sorte que son impact puisse être évincé. Le chapitre 2 fournit une nouvelle preuve que l'impact d'un serment d'honnêteté contre le mensonge dans un groupe dont les individus peuvent s'influencer mutuellement via la communication peut être évincé par une structure d'incitation qui impose une *pression sociale* sur les individus afin qu'ils se conforment au groupe. Cela met en garde les praticiens concernant les interactions entre une intervention sous serment et les incitations économiques destinées à leurs employés.

Suivant un design mixte par [Kocher et al. \(2018\)](#), l'expérience a utilisé un jeu de triche avec observation (Observed cheating game) de [Gneezy et al. \(2018\)](#). Les participants ont observé la vidéo d'un lancé de dé et son résultat. Il leur était demandé de mémoriser le résultat et de le rapporter ultérieurement pour remporter un gain. Les trois parties (intra-sujets) varient en ce qui concerne les interactions sociales. Dans la partie 1, les participants ont joué au jeu individuellement sans aucune interaction sociale. Dans la deuxième partie, les participants ont été répartis au hasard pour former un groupe de trois. Les membres du même groupe ont observé la même vidéo et ont pu communiquer anonymement pendant cinq minutes par ordinateur, après quoi ils ont rapporté *individuellement* le résultat du lancé de dé comme dans la partie 1. Enfin, dans la troisième partie, les participants ont joué au jeu dans un cadre individuel comme dans la première partie. La dimension inter-sujet a été appliquée au début de la partie 2. Les quatre traitements varient en fonction de la présence d'un serment (avec ou sans serment) et de la structure incitative des membres du groupe (avec ou sans communauté de gain ou Payoff Commonality). Le serment d'honnêteté engageait les participants à être honnêtes et à toujours dire la vérité pour le reste de l'expérience dès le début de la partie 2 (avant d'en recevoir les instructions). La règle de communauté de gain exigeait que tous les membres du groupe déclarent le même nombre pour gagner le gain. Dans le cas contraire, ils ne gagnent rien. Il y a donc quatre traitements inter-sujets : BaseNoPC, BasePC, OathNoPC et OathPC.

La principale conclusion du chapitre 2 est que le fait de prêter serment individuellement réduit le nombre de mensonges en groupe, mais l'ampleur de l'impact dépend de la structure incitative. Le serment entraîne une réduction plus importante du mensonge lorsqu'il n'y a pas de règle de communauté de gain que lorsque c'est le cas. La fraction de menteurs est plus élevée dans BaseNoPC (91,4 %) que dans OathNoPC (76,5%). En fait, l'impact du serment dans le cadre de la règle de communauté de gain est marginal (94,9% dans BaseNoPC contre 86,2% dans OathPC). En outre, la fraction de menteurs dans la partie 2 est plus élevée que dans la partie 1 pour tous les traitements, la différence étant la plus faible dans OathNoPC. Ces résultats sont en accord avec l'idée que la règle de communauté de gain peut exercer une pression sur les individus pour qu'ils se conforment au groupe, évinçant ainsi l'impact du serment à propos d'une déclaration honnête. L'analyse des comportements mensongers dans les parties 2 et 3 permet de comprendre les raisons de l'impact plus important du serment en l'absence de communauté de gain. Le raisonnement est le suivant : si les individus ont suivi les membres de leur groupe et ont menti par pure conformité dans la partie 2, on devrait s'attendre à ce qu'ils continuent à se conformer aux actes des autres lorsqu'ils sont seuls dans la partie

3. D'autre part, si certains individus ont également agi dans la partie 2 sous la pression du groupe, on pourrait s'attendre à ce qu'ils mentent moins lorsqu'ils sont seuls dans la partie 3, car l'action dans le cadre du groupe a été forcée. Alors que la pure conformité existe dans les deux structures incitatives, la pression du groupe ne devrait exister que dans le cas d'une communauté de gains.

Les données indiquent que l'effet de pure conformité est très fort dans les conditions de base. Les taux de mensonge dans un cadre individuel après les interactions de groupe sont similaires pour BaseNoPC et BasePC (91% et 92%, respectivement). Plus important encore, ces taux de mensonge ne diffèrent pas du taux obtenu dans un contexte de groupe pour le même traitement. Cela implique que la pure conformité conduit au mensonge dans un contexte de groupe dans les conditions de base, tandis que la pression du groupe sous l'effet de la communauté des gains peut jouer un rôle mineur sans serment. Cependant, il semble qu'il y ait effectivement une pression de groupe dans le cadre de la communauté de gain *avec* serment. En tenant compte de l'effet de pure conformité (telle que représentée par le nombre de menteurs dans le cadre du groupe dans la partie 2), les individus dans l'OathPC étaient environ 12 points de pourcentage moins susceptibles de mentir plus tard lorsqu'ils étaient seuls, par rapport à tous les autres traitements. Cela implique qu'un taux élevé de mensonges sous serment dans un contexte de groupe avec communauté de gain est dû à la pure conformité *et* à la pression du groupe, ce qui annule l'impact du serment. Bien qu'une préférence pour l'efficacité ne puisse être totalement exclue, il devrait coexister avec une pression des pairs poussant à mentir. Dans l'ensemble, les résultats du chapitre 2 soulignent qu'une intervention sous forme de serment peut interagir avec la structure incitative qui régit le groupe. Par conséquent, pour être efficaces dans un contexte organisationnel, les entreprises doivent mettre en œuvre une intervention sous serment en étroite relation avec les systèmes d'incitation.

Jusqu'à présent, les chapitres 1 et 2 ont montré qu'un serment d'honnêteté, en tant qu'intervention visant à augmenter le coût moral du mensonge (Abeler et al., 2019), peut décourager les comportements mensongers dans un cadre stratégique où le mensonge crée une externalité négative et lorsque les individus se trouvent dans un cadre collectif. Ils ont contribué à identifier quand un serment fonctionne et, surtout, quand son effet peut être atténué. Les résultats des chapitres 1 et 2 ont des implications pratiques importantes concernant la mise en œuvre d'un serment dans un contexte de marché et au sein d'une organisation, respectivement, afin de maximiser son effet enviable contre les comportements contraires à l'éthique.

Les résultats des chapitres 1 et 2 sont cohérents avec les cadres théoriques de l'aversion pour la tricherie perçue (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Khalmetski and Sliwka, 2019; Abeler et al., 2019) qui suggèrent que les indi-



vidus subissent un coût moral du mensonge causé par leurs préférences intrinsèques ainsi qu'un souci de réputation. Ce coût intrinsèque est intégré dans leur décision de mentir. Le chapitre 3 adopte une perspective différente des deux premiers chapitres en se concentrant sur ce qui se passe *après* que les gens ont menti. Plus précisément, une fois que les gens ont subi un coût moral lié au mensonge, le rôle de ce coût moral dans le processus décisionnel s'arrête-t-il là ? Ou continue-t-il d'influencer la façon dont les gens utilisent l'argent qu'ils ont gagné de façon *malhonnête* ? Si l'influence d'un coût moral persiste dans la manière dont les individus prennent leurs décisions économiques, utilisent-ils *l'argent malhonnête* différemment de celui gagné de manière éthique ? Le traitent-ils davantage comme un gain inattendu ou comme un gain obtenu au prix d'un effort coûteux ? Le chapitre 3 a étudié ces questions dans le contexte d'une décision économique – la manière dont les individus prennent des risques avec leur argent.

### **Chapitre 3 : Prise de risque avec de l'argent gagné de manière non éthique**

Le chapitre 3 vise à comprendre comment les gens utilisent l'argent qu'ils ont gagné par des moyens non éthiques. Plus précisément, il cherche à savoir si l'argent gagné de manière non éthique est traité davantage comme un gain inattendu que comme le résultat d'un effort coûteux dans le contexte d'une prise de risque.

Cette question se prête très bien à l'étude par des méthodes expérimentales car il est presque impossible d'observer comment les individus utilisent de l'argent gagné malhonnêtement dans le monde réel, plus particulièrement s'ils l'utilisent différemment de l'argent gagné par des moyens éthiques. Cette réflexion est liée à la comptabilité mentale (Thaler, 1985, 1999; Abeler and Marklein, 2017) qui montre que la source des gains influence la façon dont les individus dépensent leur argent. Ce type de comportement viole le principe des biens fongibles (Abeler and Marklein, 2017), qui postule que toute unité d'argent est substituable (c'est-à-dire qu'un dollar d'un compte est interchangeable avec le même dollar provenant d'un autre compte). Si les gens prennent des risques avec leur argent différemment selon qu'il a été gagné de manière éthique ou non, cela peut avoir des implications sur la prise de risque dans des domaines tels que les marchés financiers où l'exploitation de l'information asymétrique peut être tentante.

Une étude récente d'Imas et al. (2020) a exploré une violation de la fongibilité de l'argent d'origine non éthique dans le contexte des dons de charité. Dans plusieurs expériences, ils ont constaté que les individus malhonnêtes se livraient à un blanchiment d'argent mental motivé – un simple échange de billets physiques d'argent gagné de manière non éthique avec de *l'argent propre* réduit considérablement la tendance avec laquelle les individus malhonnêtes donnent leur argent à une œuvre

de bienfaisance. Le chapitre 3 diffère de celui d'[Imas et al. \(2020\)](#) sur deux points. Premièrement, il étudie une violation du principe de fongibilité dans le contexte de la prise de risque, qui n'est pas liée au domaine de la morale. Deuxièmement, [Imas et al. \(2020\)](#) ont étudié un raisonnement motivé engagé par des individus malhonnêtes pour dissocier l'argent non éthique de sa source. En revanche, en étudiant le degré de risque que les individus sont prêts à prendre avec de l'argent malhonnêtement gagné par rapport à de l'argent honnêtement gagné, le chapitre 3 explore une violation *non motivée* basée sur des mécanismes identifiés lors d'études des violations de fongibilité d'argent gagné par chance ou par effort.

En économie comportementale, la littérature a montré que les gens ont tendance à utiliser l'argent gagné au prix d'un effort coûteux différemment d'un gain inattendu (ex. [Hoffman et al., 1994](#); [Cherry et al., 2002](#); [List and Cherry, 2008](#); [Hvide et al., 2019](#)), ce qui a été expliqué par un sentiment de légitimité : un effort coûteux induit un sentiment de propriété. Suite à une prise de risque, un sentiment de légitimité devrait induire un comportement d'aversion au risque. En revanche, l'obtention d'un gain inattendu donne un faible sentiment de légitimité, ce qui entraîne une plus grande tendance à la prise de risques. La question de savoir comment les gens prennent des risques avec de l'argent gagné de manière non éthique n'est pas si simple et cela pour deux raisons.

D'une part, les gens peuvent percevoir l'argent gagné par des moyens malhonnêtes ou mensongers comme quelque chose de facilement gagné, de l'argent "house money" ([Thaler and Johnson, 1990](#)), parce qu'*objectivement*, ce n'est pas leur argent (c'est-à-dire qu'il appartient à d'autres personnes). Si tel est le cas, on peut s'attendre à une prise de risque élevée avec de l'argent gagné de manière non éthique, qui serait traité davantage comme un gain inattendu. D'autre part, les individus peuvent subir le coût moral d'avoir obtenu de l'argent de façon non éthique en raison d'une aversion intrinsèque pour le mensonge ou des problèmes de réputation associés à une aversion pour la tricherie perçue ([Dufwenberg and Dufwenberg, 2018](#); [Gneezy et al., 2018](#); [Khalmetski and Sliwka, 2019](#); [Abeler et al., 2019](#)). De tels coûts moraux pourraient s'apparenter à un coût d'effort qui induirait alors une aversion au risque. En d'autres termes, les individus malhonnêtes peuvent justifier d'un sentiment de légitimité à un gain non éthique et donc se comporter comme s'ils y avaient *réellement* droit. Ce raisonnement est basé sur l'hypothèse du coût psychologique de [Thielmann and Hilbig \(2019\)](#), selon laquelle, bien que les individus malhonnêtes n'aient pas objectivement droit au gain non éthique, ils peuvent justifier *subjectivement* qu'il *leur* appartient parce qu'ils ont encouru un coût moral pour l'obtenir. Si tel est le cas, subir un coût moral pour obtenir de l'argent peut réduire la volonté de le mettre en danger par des décisions futures.

Nous avons mené une expérience en ligne où les participants gagnent d’abord une dotation, après quoi ils sont amenés à prendre une décision risquée qui peut, en cas d’échec, réduire les gains réalisés dans la première étape. Nos trois designs inter-sujets varient selon la façon dont les participants gagnent la dotation (soit par la chance dans une loterie binaire, soit par un effort coûteux dans une tâche à effort réel, soit de façon non éthique dans une tâche de mensonge). Nous avons constaté que dans le contexte de la prise de risque, les individus traitent l’argent gagné de manière malhonnête davantage comme un gain inattendu dû à la chance que comme un gain basé sur l’effort. Cet effet est particulièrement répandu chez les menteurs ayant une aversion pour le risque.

Néanmoins, cette constatation pourrait résulter soit d’un effet de sélection des personnes ayant un faible coût moral (car le mensonge était observable a posteriori par l’expérimentateur), soit du fait qu’un coût moral ne génère aucun sentiment de légitimité. Pour distinguer ces deux possibilités, nous avons mené une expérience de suivi, en conservant le même effet de sélection mais en augmentant le coût moral (*ex-post* au mensonge) en utilisant des réponses aux normes sociales obtenues d’un échantillon distinct issu du même groupe de sujets. En particulier, les participants à l’expérience de suivi ont été exposés aux normes injonctives de la tâche de mensonge, à savoir que l’option de la vérité (ou du mensonge) était considérée comme socialement appropriée (ou inappropriée). Cette manipulation du coût moral a été effectuée après la tâche de mensonge mais avant la tâche de risque. Nous avons constaté que l’augmentation du coût moral élimine la différence entre la prise de risque avec de l’argent gagné de manière malhonnête et avec de l’argent gagné grâce à un effort coûteux. Cela implique que les coûts moraux induisent un effet de légitimité. Ces résultats ont des implications concernant le rôle du coût intrinsèque du mensonge pour limiter l’utilisation imprudente de l’argent, comme les jeux d’argent par exemple. Cela peut être pertinent dans un contexte tel que celui des marchés financiers où la fraude n’est pas rare. [Egan et al. \(2018\)](#) ont documenté que 44% des conseillers qui ont été licenciés pour fautes ont trouvé un nouvel emploi dans la même année. En outre, il existait un marché de “matching” pour les fautifs - les conseillers ayant des antécédents de faute recherchaient des entreprises qui embauchaient des personnes semblables, tandis que ces entreprises étaient moins susceptibles de licencier des conseillers pour de nouvelles fautes professionnelles. Si le coût moral de l’obtention d’argent non éthique est faible, cela implique alors une tendance à prendre des risques, ce qui peut indirectement influencer la manière dont les conseillers allouent les portefeuilles de leurs clients (ex. [Foerster et al., 2017](#); [Kling et al., 2019](#); [Linnainmaa et al., 2021](#)).

\*\*\*

En conclusion, cette thèse contribue à la littérature de l'économie comportementale de la malhonnêteté en étudiant l'influence d'un coût moral découlant de motivations intrinsèques individuelles sur la décision de mentir et si un tel coût continue d'influencer le processus de décision, même après avoir menti. Bien que nous devions encore être prudents quant à l'extrapolation des résultats de ces expériences dans un environnement plus complexe, cette thèse nous permet de mieux comprendre quand et pourquoi les gens mentent, comment concevoir de manière optimale les interventions visant à réduire la malhonnêteté au niveau individuel et collectif, ainsi que le rôle du coût moral sur l'utilisation économique de l'argent dans un domaine à risque.

Dans l'ensemble, les résultats présentés dans cette thèse fournissent une réflexion sur le processus de décision moral de *l'homo moralis*, qui résonne avec la philosophie morale d'Adam Smith. En tant que créatures sociales, nous nous comportons d'une manière qui ne maximise pas toujours notre propre avantage matériel. Nous nous soucions non seulement du bien-être des autres, mais aussi de la façon dont nos actions peuvent être en décalage avec les règles normatives et nos motivations morales innées. Nos motivations intrinsèques *et* les croyances partagées avec les autres membres de la société constituent la boussole morale qui nous guide vers des comportements socialement adaptés et appropriés, nous permettant de vivre en harmonie les uns avec les autres. Néanmoins, comme Smith l'a souligné à propos de la nature humaine de *l'amour de soi*, nos passions peuvent nous mener à notre perte, nous aveuglant et nous faisant agir en désaccord avec les règles et devoirs moraux. Sommes-nous capables de réfréner nos désirs égoïstes et de promouvoir une amélioration socio-économique de l'humanité ? Ou sommes-nous perdus parce que nous ignorons notre boussole morale ?

# Contents

<b>Acknowledgements</b>	<b>ii</b>
<b>Résumé de la thèse</b>	<b>iv</b>
<b>General Introduction</b>	<b>2</b>
<b>1 Honesty Oaths in Advisor-Client Relationships: An Experiment</b>	<b>32</b>
1.1 Introduction . . . . .	32
1.2 Related Literature . . . . .	36
1.3 Experimental Design and Procedures . . . . .	38
1.4 Conjectures . . . . .	42
1.5 Results . . . . .	44
1.5.1 The Impact of a Common-Knowledge Oath on Advisors and Investors . . . . .	45
1.5.2 Understanding Investors' Behavior in the Oath-CK Treatment	49
1.5.3 Lying under a No-Common-Knowledge Oath . . . . .	51
1.5.4 The Effect of Oath Over Time . . . . .	54
1.6 Discussion and Conclusion . . . . .	56
Appendix . . . . .	59
1.A Instructions (Translated from French) . . . . .	59
1.B Appendix Tables . . . . .	67
1.C Appendix Figures . . . . .	71
1.D Belief Elicitations . . . . .	74
1.E Survey on Prolific . . . . .	78

---

<b>2</b>	<b>Individual Oath-Swearing and Lying in Group</b>	<b>82</b>
2.1	Introduction . . . . .	82
2.2	Related Literature . . . . .	86
2.3	Experimental Design and Procedures . . . . .	88
2.4	Conjectures . . . . .	91
2.5	Results . . . . .	93
2.5.1	The Impact of Oath on Lying in Group . . . . .	93
2.5.2	Group Pressure under Payoff Commonality . . . . .	97
2.5.3	Chat Analysis . . . . .	100
2.6	Discussion and Conclusion . . . . .	105
	Appendix . . . . .	108
2.A	Instructions (Translated from French) . . . . .	108
2.B	Appendix Tables . . . . .	116
2.C	Appendix Figures . . . . .	117
2.D	Code Book for Content Analysis (Translated from French) . . . . .	118
2.E	Additional Analyses . . . . .	120
<b>3</b>	<b>Risk-Taking with Unethically Earned Money</b>	<b>128</b>
3.1	Introduction . . . . .	128
3.2	Related Literature . . . . .	132
3.3	Experimental Design and Procedures . . . . .	135
3.4	Conjectures . . . . .	138
3.5	Results of the Main Experiment . . . . .	140
3.6	Follow-up Experiment . . . . .	143
3.7	Discussion and Conclusion . . . . .	147
	Appendix . . . . .	151
3.A	Instructions . . . . .	151
3.B	Appendix Tables . . . . .	170
3.C	Appendix Figures . . . . .	172
3.D	Analysis of the \$1 Group Behavior . . . . .	174

---

<b>General Conclusion</b>	<b>178</b>
<b>Bibliography</b>	<b>184</b>

# List of Figures

1.C.1	Fraction of Detectable Lies in Each Period, by Treatment . . . . .	71
1.C.2	Fraction of Deniable Lies in Each Period, by Treatment . . . . .	71
1.C.3	Cumulative Distribution Function of Advisors' Earnings, by Treatment	72
1.C.4	Cumulative Distribution Function of Investors' Earnings, by Treatment	72
1.C.5	Cumulative Distribution Function of Joint Earnings of Advisors and Investors, by Treatment . . . . .	73
2.1	Distribution of Reported Points, by Treatment. . . . .	95
2.2	Frequencies of Arguments Related to Money and Morality . . . . .	103
2.A.1	Sample Video of Die Outcome . . . . .	114
2.A.2	Decision Screen in Part 1 . . . . .	114
2.A.3	Decision Screen in Part 2 . . . . .	115
2.A.4	Sample Chat Stage in Part 2 . . . . .	115
2.C.1	Frequencies of Each Type of Arguments . . . . .	117
2.E.1	Belief about Group (Dis)honesty by Type and Treatment . . . . .	122
2.E.2	Change in Beliefs about Individual Honesty After Group Interactions	123
3.1	CDF of the Number of Boxes Collected in the BRET, by Treatment .	141
3.2	Average Marginal Effects of the Treatment Dummies on Risk Taking, Conditional on Baseline Risk Attitudes . . . . .	148
3.C.1	Higher Moral Cost Treatment: Information on the Social Norm Re- garding the Honest Option . . . . .	172
3.C.2	Higher Moral Cost Treatment: Information on the Social Norm Re- garding the Dishonest Option . . . . .	172



---

3.C.3 Cumulative Distribution Function of the number of boxes collected in  
the BRET, by treatment (including the Higher Moral Cost treatment) 173

# List of Tables

1.1	Detectable and Deniable Lies in the No Oath and Oath-CK Treatments . . . . .	46
1.2	Investments in the Announcement and Truthful Announcement Games, by Treatment . . . . .	48
1.3	Investment Rates in the Oath-CK and Oath-NoCK-Async treatments	50
1.4	Detectable and Deniable Lies in the Oath-CK and Oath-NoCK Treatments . . . . .	52
1.5	Probability of Making Deniable Lies . . . . .	55
1.B.1	Summary Statistics: All Participants . . . . .	67
1.B.2	Summary Statistics: Advisors . . . . .	67
1.B.3	Summary Statistics: Investors . . . . .	67
1.B.4	Investments in the Announcement and Truthful Announcement Games, by Treatment . . . . .	68
1.B.5	Average Earnings of Advisors and Investors, by Treatment . . . . .	68
1.B.6	Probability of Making Detectable Lies . . . . .	69
1.B.7	Probability of Investing in 2 and 3 Stars Announcements . . . . .	70
1.D.1	First Order Belief of Investors . . . . .	74
1.D.2	Distance Between Beliefs of Investors and Actual Honesty . . . . .	75
1.D.3	First Order Belief of Advisors . . . . .	76
1.D.4	Second Order Belief of Advisors . . . . .	76
1.D.5	Distance Between Beliefs of Advisors and Actual Honesty . . . . .	77
2.1	Lying under Oath in a Group Setting . . . . .	96
2.2	Individual Lying in Part 3: Pressure vs. Conformity . . . . .	99
2.3	Chat Duration and Volume of Messages . . . . .	101

---

2.4	Impact of Arguments on Lying in Part 2 . . . . .	104
2.B.1	Summary Statistics . . . . .	116
2.B.2	Number of Groups by Types of Arguments . . . . .	116
2.B.3	Shares of Dishonest and Honest Arguments . . . . .	116
2.E.1	Coordination Among Group Members by Treatment . . . . .	120
2.E.2	Response Time (in seconds) in Part 2 . . . . .	125
3.1	Determinants of Risk Taking in the BRET in the \$2 Group . . . . .	142
3.2	Determinants of Risk Taking in the BRET in the \$2 Group (Original and Follow-up Experiments) . . . . .	146
3.B.1	Summary Statistics (All participants) . . . . .	170
3.B.2	Summary Statistics (\$2 group) . . . . .	170
3.B.3	Summary Statistics (\$1 group) . . . . .	171
3.B.4	Determinants of the Probability of Earning \$2 in Part 1, by Treatment	171
3.D.1	Determinants of Risk Taking in the BRET in the \$1 group . . . . .	174
3.D.2	Windfall Treatment: \$1 <i>vs.</i> \$2 group . . . . .	176
3.D.3	Effort Treatment: \$1 <i>vs.</i> \$2 group . . . . .	176
3.D.4	Lying and Higher Moral Cost treatments: \$1 <i>vs.</i> \$2 group . . . . .	177

# General Introduction

*“To feel much for others and little for ourselves, that to restrain our selfish, and to indulge our benevolent affections, constitutes the perfection of human nature ...”*

— Adam Smith, *The Theory of Moral Sentiments*

Morality is incorporated in all of us. We have a moral compass that guides our actions and warns us when we are *morally* astray. Does our moral compass function so perfectly that it steers us away from *all* immoral behaviors? Do we always follow it, or we sometimes turn a blind eye to it, voluntarily or not? The answers to these questions may not be that simple.

Our moral compass often guides us in an accurate direction. We are aware of the social norms – what kinds of behavior do most people consider as morally appropriate and inappropriate. Despite of having a shared belief that prescribes the unwritten social rules, we *do* violate them at times. For instance, lying is deemed as an inappropriate thing to do, yet people still lie. Is it because our moral compass considers that some lies are *less* immoral than the others? Or is it just *us* who ignore the rules by trying to justify certain kinds of lies?

One may argue that the degree of immorality depends on the motives behind those lies. Benevolent white lies, meant to reduce social frictions, are often perceived as less problematic (and therefore less unacceptable) because one may think that their benefit outweighs their cost. This ranges from how we sugarcoat the truth from our loved ones (like the times you tell your partner that the color of the outfit complements his or her skin) to how parents tell children to behave themselves before Christmas (else Santa Claus will not come bearing gifts). Beyond these seemingly-harmless lies are those self-serving ones that bring personal benefits at the cost of others. We are confronted in our daily lives with opportunities to enrich ourselves through unethical means – be it small, like when we find a lost wallet on the street, or larger ones with more serious consequences like tax evasion and embezzlement. It is such misbehaviors that our moral compass tries to steer us away from. Yet, we ignore it at times, consciously or not, trying to seek justifications in favor of an

unethical pursuit, while still feeling at peace with our own conscience.

We are the product of our own evolution, and so is our sense of what is right and wrong. As societies progress through the passage of time, individuals learn about the normative rules. Actions that bring social beneficence are rewarded, while those that bring social maleficence are sanctioned by others. This leads to a construct of normative standards where righteous acts are reinforced and continued, and immoral ones are suppressed and discontinued. Individuals then internalize these social rules, and as they produce offsprings, these virtues are passed on from generation to generation. The entire process continues, shaping the sets of normative rules in our present societies. In short, the moral compass (our sense of morality) is a product of a biological and cultural co-evolution (Gintis et al., 2005; Gintis and Helbing, 2015).

Despite of having internalized the normative standards, we can sometimes feel tempted to break them and behave in a self-serving manner. Doing so creates a cognitive dissonance within us because there is a misalignment between our sense of morality and our own demeanor. We may feel internally disturbed from the negative emotions such as guilt and shame because we are aware that we have deviated from what is considered a morally fit behavior. These *psychic costs* can be large when our misdeed poses harms to others. Further, they can be amplified when we are exposed to the judgement of others because they expect us to abide by the rules, and we know it. Therefore, when individuals violate the normative rules, which they have themselves constructed and internalized, they experience *moral costs* arising from the breach of their intrinsic motivations.

## Objectives

This thesis lies at the intersection of morality and economics. It contributes to the constantly flourishing field of behavioral economics of dishonesty by providing novel evidence on the influence of moral costs on lying and economic decisions. Relying on the methodology of experimental economics, it aims test the extent to which the moral costs, arising from individual intrinsic motivations, can limit lying behavior, under what circumstances their influence may be limited, and whether they can affect the way people make an economic decision.

In brief, Chapter 1 and 2 focus on understanding the influence of a moral cost in limiting lying behaviors in the contexts of a financial market setting and a group setting, respectively. Chapter 3 takes another perspective by inquiring into the persistence of a moral cost on the decision-making process *after* one has lied. In essence, it aims to understand whether the moral cost of lying influences the way

people use their unethically earned money.

More precisely,

- Chapter 1 investigates the impact of an honesty oath against lying behavior in a financial market setting using a laboratory experiment.
- Chapter 2 examines the impact of an individual oath-swearing against lying in a group setting using a laboratory experiment.
- Chapter 3 investigates whether people treat dishonest money earned at a moral cost more like a windfall gain or as the result of costly effort in the context of risk-taking using an online experiment.

The remainder of this general introduction proceeds as follows. It first encapsulates the moral philosophies of Adam Smith and Immanuel Kant and their relevance to the field of modern economics. Then, it depicts how economists traditionally viewed the issue of dishonesty and how the approach has evolved with behavioral economics, whose considerations very much echo with the ideologies of Smith and Kant. The final portion describes the methods behavioral economists use to study dishonesty, the empirical evidence from the literature and each of the chapters with its novel contributions in detail.

## **Morality and Economics**

Adam Smith's *Theory of Moral Sentiments*, published in 1759 prior to the widely-known *Wealth of the Nations*, shed light on why morality matters for economics. In the *Theory of Moral Sentiments*, [Smith](#) asked two questions about the moral foundation of human society – What is *virtue* and how does it come about?

Smith argued that we learn about morality in two ways. The first is through the concept of *sympathy*, which is closer to a contemporary meaning of the word 'empathy.' It is the ability to imagine oneself in the shoes of the other person and feel what he or she feels. Through *mutual sympathy* (or fellow-feeling), one learns whether the action is appropriate or inappropriate. This is stated in the opening sentence of the *Theory of Moral Sentiments*:

*“However selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.”* — (*Theory of Moral Sentiments*, First.I.I., p.3)

In contemporary language, Smith stated that our human capacity to sympathize with each other allows us to learn context-dependent rules of conduct, eventually enabling us to live harmoniously as a community. Sympathy allows each individual to observe the actions and reactions of others and learn about what brings *praiseworthiness* and *blameworthiness*. Using our initial example, keeping a lost wallet and not trying to return it brings about disapproval of others. By observing those *disapprobations*, we can learn to adjust our behavior to be consistent with what is socially approved. In short, it is the desire to be approved by others that drives us to behave in congruence with others in the society.

Smith also proposed the concept of an *impartial spectator* as a second way through which we learn moral conducts. It serves as an *internal* rule that guides our moral compass. Similar to the contemporary meaning of the word ‘conscience’, it is the ability to project oneself as a *disinterested cold* observer of one’s action who makes a moral judgement. This serves to overcome the prejudice others may have when judging our action and as a substitute through which we learn from others in the society. Returning to the lost wallet example, Smith’s impartial spectator would work in the following way – you have found a lost wallet. You can either keep it or try to find its owner. Before making the decision, you imagine yourself judging each possible action. Would your *imagined* self, free from all the biases and emotions, approve or disapprove of the action to be taken? In summary, Smith argued that it is through our mutual sympathy and the impartial spectator that we learn the *norm of propriety* – what is expected of us and what is seen as ‘fit and proper.’

Nonetheless, Smith acknowledged the limitations of the impartial spectator. Smith remarked the very innate nature of *self-love* in all of us.

*“Every man is, no doubt, by nature, first and principally recommended to his own care; and as he is fitter to take care of himself than of any other person, it is fit and right that it should be so.”* — (Theory of Moral Sentiments, Second.II.II., p.119)

The sense of self-love, which can be non-satiated and arrogant, can lead our *passions* to blind us in the heat of the moment when we are about to take an action, or even afterwards because we might perceive the rules differently when contemplating upon our past immoral action. Smith referred to this justification process as ‘*self-deceit*.’ In the lost wallet scenario, one may fall prey to own greed, keep the wallet and even justify ex-post the selfish decision. These failings result in a misjudgement at an individual level and the society at large. Because it is not easy to avoid making such errors in moral judgement, Smith branded self-deceit as a “*fatal weakness of mankind*” and “*the source of half the disorders of human life*” (Theory of Moral Sentiments, Third.(I).III, p. 222).

Overall, Smith's Theory of Moral Sentiments provided a moral foundation of *human social betterment* – How can we, as social creatures, prosper and live with one another in harmony? Smith's ideas are very much concurrent with the works on social norms (see [Elster, 1989](#); [Ostrom, 2000](#); [Bicchieri, 2005](#)). We learn about norms from others, and vice versa. Morally fit actions are reinforced by social approval while those deemed morally unfit are dissuaded by social sanctions. We imbibe these rules into our moral conscience that automatically guides us, telling us what is right or wrong. We do not try to only pursue what is best for ourselves but also to meet the expectations of others, governed by a set of legal and social rules – both written and unwritten. However, we may *give in* to the pursuit of *shameless self-promotion*, leading us to find self-serving justifications to validate our questionably improper actions before attempting or even after having done so.

Another ethical concept that speaks to the foundation of human morality, similarly to that of Smith, is Kantian ethics – a deontological moral philosophy of Immanuel Kant, laid out in his work entitled '*Groundwork of the Metaphysic of Morals*' in 1785. While [Kant](#)'s moral principle also relates to social norms as in the case of Smith's, it slightly differs in that everyone, in a Kantian world, knows and follows the moral rules, which do not arise from social learning but from one's *rationality* and *intellect*.

At the core of Kantian ethics lies the *categorical imperative*, which prescribes behaviors under the *moral rules* and *duties* for which everyone in the society is bound to follow and uphold, *regardless* of one's desires. According to Kant, these moral obligations are the outcomes of *rational* thoughts – one only needs pure reasoning and intellects to reach a conclusion regarding the moral appropriateness of an action. Thus, Kant's categorical imperative is a set of moral rules that is absolute – either the action is moral or immoral. There is no middle ground to immorality and it applies equally to all individuals.

Kant posited that the moral rules are based on the *principle of universality* – Can the action in question become universalized? More precisely, Kant stated to “*act only according to that maxim which you can at the same time will that it should become a universal law without contradiction.*” In contemporary language, the action is moral *if and only if* it can be generalized to everyone because it is within the universal rule that permits everyone to do the same. On the other hand, if the action cannot be universalized, it is not a moral act.

When a Kantian considers whether it is right or wrong to steal, he or she will first question its maxim (*i.e.*, what is the rule behind stealing?) then reason whether it should be permitted for every individual in the society (*i.e.*, should everyone steal and what are the collective outcome if everyone steals). By this point, if one thinks



that stealing is not universalizable, then it is not allowed under the universal moral rules and doing so leads to a contradiction with those rules.

Besides the universality principle, Kant also formulated an ethical principle that relates to *humanity*, which demands proper treatment towards others. More precisely, Kant stated to “*act so that you treat humanity, whether in your own person or in that of another, always as an end, and never as mere means.*” Simply put, people should not be exploited and used to their disadvantage. By manipulating others, we rob them of their *autonomy*, which is an individual entitlement. By this formulation, Kant prescribed that humanity should treat one another with respect and that we should recognize others as *ends* rather than *means*.

If one views a typical economic transaction from a Kantian perspective, when a salesman is tempted to cheat his customer by recommending a low-quality product as a *high-quality* one (and perhaps charge a high price for it), he will not do so *unless* he believes that *all the salesmen* should do the same. In addition, by deceiving the customer about the true quality of the product, he or she is deprived of the decision right (*i.e.*, the ability to freely decide, knowing that the product is in fact of low-quality). Therefore, the salesman in the Kantian world will not cheat the customer because the action is not permitted by the universal moral rules and that it violates the client’s autonomy.

To reiterate, similar to Smith’s moral philosophy, the notions of Kantian ethics are analogous with how individuals care about social and moral norms. Our actions are guided by the injunctive norms (*i.e.*, what one ought to do) and we are aware of them. We try to remain within the boundary of these universal rules by avoiding behaviors others would deem as inappropriate. Because we live in a society with normative rules, we have developed *other-regarding preferences* and *intrinsic motivations* to behave morally, besides our desires to achieve personal goals (Gintis et al., 2005). These ideas correspond to the behavioral economic perspective of human decision-makings. We care about the welfare of others – we treat other people fairly (*e.g.*, Rabin, 1993; Fehr and Schmidt, 1999; Charness and Rabin, 2002; Falk et al., 2003), pay attention to how an outcome is achieved (*e.g.*, Frey et al., 2004; Falk and Fischbacher, 2006; Falk et al., 2008), reciprocate kindness with kindness, and punish those who transgress us, or even others (*e.g.*, Fehr and Gächter, 1998; Abbink et al., 2000; Fehr and Gächter, 2000; Masclet et al., 2003; Fehr and Fischbacher, 2004). Besides these other-regarding preferences, we are motivated intrinsically to treat people with decency and respect with any malignant intent, just because we simply think *it is the right thing to do* (Gintis, 2016). These innate motives encompass virtuous characteristics such as honesty, loyalty, trustworthiness, diligence and courage. It is pursued, at least in part, for *its own sake* rather than for *its effect*.

The sense of morality, as Herbert Gintis puts it, is a product of biological and cultural co-evolution (Gintis and Helbing, 2015). It is an outcome of genetic evolution because normative rules are transmitted from generation to generation. One can imagine how societies transmit socially desirable traits to their off-springs (see Kosse et al., 2020; Abeler et al., 2021), while immoral traits are sanctioned and discontinued (see Balafoutas and Nikiforakis, 2012; Balafoutas et al., 2014; Brouwer et al., 2022). In addition, this evolutionary process depends on the cultural context, giving rise to heterogeneous normative rules across societies (Henrich et al., 2004).<sup>1</sup>

Nonetheless, we may need to trade off these intrinsic motivations (what Gintis refers to as a universalist motive of an action) with self- or other-regarding ones, depending on how much the (im)moral action costs to oneself, and brings harms or benefits to the others. Such a trade-off lessens the value of honesty. Using our previous example of a salesman, it is possible that the salesman may not be able to afford honesty given his personal financial situation, thereby causing him to deceive the customer. On the other hand, if the salesman recognizes that deceiving the customer will pose great harm (say using a low-quality product can endanger the customer's welfare), then the likelihood to defraud is reduced. Thus, even though we are guided by the normative rules, care for others in the community, and have innate desires to be moral, we still face moral dilemma sometimes, depending on the circumstances. An important point, however, is that *at least* we try to be *moral*. Thus, the way we make decisions in a moral domain is more like a *homo moralis* (Gintis and Helbing, 2015) – a man who has an innate desire to act morally – rather than a *homo economicus* – a man who bases his decision on a mere calculation of costs and benefits.

### **An Economic Approach to Studying Dishonesty**

The way traditional economists thought about (im)moral behaviors were far from the ideologies of Smith and Kant. Economists were first interested in criminal behaviors because of the economic consequences of criminal offenses on the society. The seminal work of Gary Becker ‘*Crime and Punishment: An Economic Approach*’ in 1968 pioneered the theoretical model for economics of crime.

The model of Becker (1968) proposes that the decision to commit a crime is an outcome of a cost-benefit analysis. That is, when a person confronts with an opportunity to commit a criminal offense, he or she calculates the benefit (*i.e.*, monetary gains from theft) and the cost (*i.e.*, expected cost of punishment given the probability of getting caught). Mathematically, the decision maker  $i$  in Becker's

---

<sup>1</sup>See also papers by Henrich et al. (2001, 2006), Marlowe et al. (2008), Falk et al. (2018) and Cohn et al. (2019) on the heterogeneity in preferences across cultures.

basic framework derives his or her expected utility from committing a crime as follows.

$$EU_i = p_i U_i(Y_i - f_i) + (1 - p_i) U_i(Y_i) \quad (1)$$

where  $Y_i$  represents income from the crime (*i.e.*, monetary and psychic),  $f_i$  and  $p_i$  represent the monetary punishment and the probability of getting caught, respectively. In short, in Becker's framework, a choice to commit a crime is solely based on a calculation of the economic benefit it brings to the table at a possible cost of legal sanctions. Therefore, a *rational* decision maker in Becker (1968) will commit a crime if the marginal benefit exceeds the marginal cost, otherwise he or she will not.

Becker's model can explain several criminal behaviors— be it severe or not. For instance, one can think about illegal car parking. The driver can either park a car in a forbidden area but remains relatively close to the destination or park it in an authorized space, which can be quite far from the destination. In Becker's framework, the benefit of illegal parking would be saving resources (*i.e.*, time, effort, money, gas) from searching for a parking space. On the other hand, this comes at a cost of getting sanctioned by the legal authority. Therefore, in such a situation, the driver would calculate whether it is *worth it* to park illegally – How much resources can be saved by parking in a forbidden area? What are the chances that the police will pass by, and in that case, how much fine will have to be paid? In a similar vein, a rational bribe taker considers the monetary benefit from accepting a bribe and the potential cost arising from engaging in corruption (*i.e.*, the probability of getting caught and punished such as getting discharged from work or imprisoned by the legal authority). Becker's model thus emphasizes on the role of audit probability and severity of punishment as *deterrence incentives* against criminal behaviors. Applying Becker's framework to dishonest behaviors of laymen (for instance lying behavior), it would imply that a rational person will tell a lie (*i.e.*, misreporting income tax, defrauding customers) if financial gains from lying outweigh the associated costs arising from getting caught in expectation.

The theoretical framework of Becker (1968) initiated the interests in the field of the economics of crimes.<sup>2</sup> Although it does not reject the role of *intrinsic motivations* as the determinants of unethical decision-makings, it applies a cost-benefit analysis, which offers a rather *cold* perspective of (im)moral behaviors. Studies on tax evasion have demonstrated that individuals indeed react to the changes in the probability of

---

<sup>2</sup>See Allingham and Sandmo (1972), Becker and Stigler (1974) and Polinsky and Shavell (1979) for early theoretical works following Becker's approach.

detection and the severity of sanction (*e.g.*, Alm et al., 1992, 1993; Slemrod et al., 2001).<sup>3</sup> However, the standard economic-of-crime models fail to explain certain behavioral patterns in individual cheating behavior. First, even in the absence of any detection, there exists heterogeneity in individual's moral behavior. Notably, cheating behavior is usually lower than what the standard economic model would predict. In fact, some individuals simply do not cheat. This means that people do not treat unethical decision like a gamble and the hesitance to misbehave reveals the existence of *moral costs*. Second, there exists incomplete dishonesty, whereby people *do not* cheat to the full extent, but rather do so *partially*. This suggests the non-linearity of moral costs and that people may care about how they are perceived by others. These two behavioral patterns cannot be explained by the standard economic model (Abeler et al., 2019), which illustrates the need to extend economic modelling to incorporate *moral* and *social* preferences into individual utility function, which lie the contributions of behavioral economists.

In fact, Amartya Sen provided several remarks on the need to broaden the notion of self-interest.<sup>4</sup> Particularly, Sen referred to Smith's sympathy, where one needs to relax '*self-goal choice*' (that is a self-interested one) in order to allow for '*the consideration of the goals of the others*' (Sen, 2004). Sen further noted that by forgoing the self-goal choice, one needs commitment that '*arises from self-imposed restriction on the pursuit of one's own goals (in favor of, say, following particular rules of conduct)*.' Parallel to the notion of *homo moralis*, Sen pointed out that people do not only care about the outcomes of their actions but rather also *how* those actions are viewed under a set of legal, social and moral rules, and by adhering to these rules, they may need to sacrifice personal goals.<sup>5</sup>

This circles back to the ideologies of Smith and Kant that individuals have innate preferences to behave in accordance with what the society prescribes. People know what are considered morally (in)appropriate and they strive to abide with them. Whenever their actions are in discord with the normative rules, they suffer not only legal *and* social sanctions but also the *moral costs* that may arise from the misalignment of the immoral action and their intrinsic preferences. These *non-monetary* (or psychological) costs are incorporated into the utility function as a *disutility* to the decision maker.<sup>6</sup> In other words, the economic perspective on moral

---

<sup>3</sup>See also Alm and Malézieux (2021) on the meta-analysis of the tax evasion game.

<sup>4</sup>See Sen (1973, 1977, 1994).

<sup>5</sup>See Vanberg (2008b) for a discussion about incorporating moral preferences into economic model. He referred to Sen's point as a move away from preferences over *outcomes* towards preferences over *actions*. See also Irlenbusch and Villeval (2015) for a discussion on behavioral economics of ethics.

<sup>6</sup>For examples, see Elster (1996, 1998) who discussed about incorporating emotions as psychic

decision-makings shifted from a purely rational one of a *homo economicus* – a man who only cares about himself – to that of a *homo moralis* – an innately morally driven man who cares not only for himself but also for others.

In any economic transaction, individuals can exploit information asymmetry to their own advantage by misrepresenting, or simply withholding, the truth at a cost to another party. For instance, when getting a car repaired, the mechanic may defraud his customer by not performing a faithful diagnosis, not resolving all problems or simply charging for services that may be unnecessary. The customer, on the other hand, may not be able to verify *ex-post* if all of the problems have indeed been correctly treated. If it is truly the case that the customer cannot verify *ex-post* the treatment provided by the mechanic, a *homo economicus* would always defraud the customers because doing so maximizes monetary benefits and there is no risk of detection. How would a *homo moralis* behave?

A way through which behavioral economists incorporate preferences into the utility function is through *belief-dependent* motivations. That is, while trying to maximize the monetary benefit, individuals *also* care about the beliefs of others (see e.g., Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007).<sup>7</sup> The guilt aversion model of Battigalli and Dufwenberg (2007) formally proposes a belief-dependent utility function where the decision maker cares not only the material payoff from the action but also how much he or she falls below the expectation of others. Formally, the basic two-person guilt aversion model of Battigalli and Dufwenberg (2007) can be summarized as follows.

$$u_i(z, \alpha_j) = \pi_i(z) - \theta_{ij} \max\{0, E_{\alpha_j}[\pi_j] - \pi_j(z)\} \quad (2)$$

The model in Equation 2 states that the decision maker  $i$  derives his or her utility from the material payoff  $\pi_i$  realized by the chosen action (and thereby realizing the outcome  $z$ ) and experiences guilt defined by the difference between  $j$ 's expectation of how  $i$  should decide (that is, depending on the prior belief  $\alpha_j$ ) and the actual decision of  $i$ , weighted by the guilt sensitivity  $\theta_{ij}$ . When the guilt component (the second term) is positive, the decision maker  $i$  experiences disutility from *disappointing* the person  $j$  because the action falls short of his or her expectation. The more sensitive the decision maker is to guilt (as captured by  $\theta_{ij}$ ), the more he or she tries to avoid disappointing the other person.

---

costs or benefits into the utility function under a rational choice framework. See also experimental works on the emotional costs by Coricelli et al. (2010, 2014) in the context of tax evasion and by Jaber-López et al. (2014) in the context of bribery.

<sup>7</sup>See Ellingsen and Johannesson (2004), Vanberg (2008a) and Ellingsen et al. (2010) for experimental works in this domain.

Looking at lying behavior from the perspective of a guilt aversion model, if the decision maker anticipates that the other person expects the truth from him or her, one avoids telling a lie because one wants to live up to the expectation of the other person. Thus, in the earlier example of a car mechanic, even though he earns more from defrauding his customer, he avoids feeling guilty from disappointing the customer and therefore does not cheat.

Another type of model relates to social norms by incorporating *norm-dependent* preferences into the utility function (see e.g., Kessler and Leider, 2012; Krupka and Weber, 2013; Kimbrough and Vostroknutov, 2016; Krupka et al., 2017). In the simple utility framework of Krupka and Weber (2013), the decision maker cares about the monetary benefit of the action *and* the degree to which the action is deemed as socially or morally (in)appropriate. Formally, Krupka and Weber (2013) defined the utility function as follows.

$$u(a_k) = V(\pi(a_k)) + \gamma N(a_k) \quad (3)$$

Equation 3 represents a utility function of the decision maker who is confronted with  $k$  possible actions. The function  $V(a_k)$  represents the utility derived from the material payoff from undertaking an action  $k$ , while  $N(a_k)$  represents the social norm component of the utility function, weighted by the degree to which he or she cares about the norms ( $\gamma$ ). Individuals who do not care about the social norms ( $\gamma = 0$ ) will undertake the action that yields maximum material payoff. On the other hand, those who care about the norms ( $\gamma > 0$ ) gain utility from adhering to what the society deems as appropriate ( $N(a_k) > 0$ ) and suffer disutility when the action is deemed as inappropriate ( $N(a_k) < 0$ ). The greater the decision maker cares about the norm (as captured by  $\gamma$ ), the greater utility gained or lost.

Applying the social norm framework to lying behavior – if the decision maker believes that lying is deemed as socially inappropriate, he or she suffers a disutility from telling a lie as the action deviates from the social norm that prescribes honesty. Thus, if the car mechanic in our example believes that most people (*i.e.*, mechanics alike, or even laypersons) deem defrauding as socially inappropriate, he is discouraged to cheat his customers.

Lastly, behavioral economists have developed theoretical models that incorporate *intrinsic* motivations for honesty into the utility function to explain lying behavior. Early theoretical works focused on a direct cost of lying (see e.g., Ellingsen and Johannesson, 2004; Kartik, 2009).<sup>8</sup> A recent work by Abeler et al. (2019) models

---

<sup>8</sup>Also see Kartik et al. (2007), Matsushima (2008), Ellingsen and Östling (2010) and Kartik et al. (2014) for theoretical works that accounted for some preferences for truth-telling.

the moral costs of lying through preference for *being honest* and being *seen as honest* (also see Gneezy et al. (2018), Dufwenberg and Dufwenberg (2018) and Khalmetski and Sliwka (2019) who proposed similar models). Formally, the model of Abeler et al. (2019) represents the utility function as follows.

$$\phi(r, c(r, \omega), \Lambda(r); \theta^{LC}, \theta^{RH}) = u(r) - \theta^{LC}c(r, \omega) - \theta^{RH}v(\Lambda(r)) \quad (4)$$

Equation 4 represents the decision maker's utility function  $\phi$  defined by the material payoff  $u(r)$  (the first term) and the moral costs of lying  $\theta^{LC}c(\cdot)$  and  $\theta^{RH}v(\cdot)$  (the second and third terms). The decision maker's material payoff is determined by what the decision maker states ( $r$ ), which can differ from the *true* state of the world ( $\omega$ ). Given that  $\omega$  is a private information to the decision maker, he or she can misrepresent it by stating  $r$  that is different from  $\omega$ , and by doing so typically derives greater monetary benefit. In the car mechanic example, he knows what is the truth, represented by  $\omega$  (*i.e.*, whether he has correctly treated the problem at a fair price), but can lie to the customer, represented by  $r$  (*i.e.*, 'claimed' to have treated in good faith at a 'fair' price). Thus, the car mechanic can earn more material payoff by lying to the customer about the services provided and the price ( $r \neq \omega$ ).

The first component of the moral costs of lying can be thought of as the *pure* lying cost. Depending on the decision maker's type (as captured by  $\theta^{LC}$ ), he or she suffers some psychological costs whenever  $r \neq \omega$ . This component of lying cost models *preference for being honest* – a person experiences disutility when he or she lies and this can come from a self-image concern (see *e.g.*, Mazar et al., 2008; Barron, 2019; Barron et al., 2022) that may arise from various reasons such as religious beliefs, internalized norms and values. In the car mechanic example, when he defrauds the customer, he incurs the moral cost of lying depending on his moral type, just because he *lies*. The greater he cares about his own preference for honesty (the higher  $\theta^{LC}$  is), the greater will his disutility be from lying.

The second component of the moral costs of lying is usually referred to *reputation* cost. Depending on the decision maker's type (as captured by  $\theta^{RH}$ ), he or she suffers some psychological costs, depending on the observed action  $\Lambda(r)$ . Simply put, the reputation cost captures the disutility arising from what other people *perceive* him or her to be (*i.e.*, an honest or a dishonest type) based on what they actually observe about him or her (*i.e.*, what the decision maker has stated). Thus, when our car mechanic lies to the customer, he takes into account how he is perceived in the eyes of the customer. Does the action make him look like an honest tradesman or more like a fraudster? If he thinks that what he would tell (*i.e.*, what the customer actually observes) would make him look like a fraud *and* he is sensitive to how others

might perceive him (the higher  $\theta^{RH}$  is), then he suffers a high disutility in terms of *reputation*, which discourages him from lying.<sup>9</sup> In summary, the model of [Abeler et al. \(2019\)](#) states that people care about being honest as well as being *perceived* by others as an honest person. Together, these two preferences constitute the *moral cost* of lying.

### Empirical Evidence in Behavioral Economics of Dishonesty

Behavioral economists provide empirical evidence by relying on experimental methods. In general, economic experiments, in particular those conducted in a laboratory setting, allow for a rigorous testing of theoretical frameworks under a controlled, anonymous and incentivized environment. It allows behavioral economists to introduce an exogenous variation and observe the impact of that variation on the outcome variable of interest, *ceteris paribus*.

Experimental method is especially beneficial for research on dishonest behavior. First, dishonesty is usually hidden. Second, when one uses natural data for dishonest behaviors that can be observed (*i.e.*, criminal activities), they do not represent the whole picture as data exist only for those who have been caught. Lastly, it can be challenging to disentangle mechanisms and identify the true drivers of dishonest behaviors in the field. Thus, experimental economics alleviates these problems by using convenient samples (typically university students or workers on crowd-sourcing platforms) who are put in a *scenario* where their decisions have real impacts – they can earn money from their decision-makings and sometimes their actions can hurt others. Notwithstanding these advantages, there are natural limitations to a laboratory experiment. One can be concerned about the external validity of the findings from a controlled setting like a lab or an online experiment as the context can be seen as *artificial* when compared to the real world. Yet, lab studies, when carefully designed and executed, do have internal validity and can still tell something about the complex world because they can help identify mechanisms driving behaviors.<sup>10</sup>

Two types of lying tasks are typically used to investigate dishonest behavior.<sup>11</sup> The first task is a two-player communication game which models lying behavior with strategic interactions. This task is commonly referred to as a Sender-Receiver Game or Deception Game, introduced by [Gneezy \(2005\)](#) with a theoretical foundation in

---

<sup>9</sup>There is a slight difference here between [Dufwenberg and Dufwenberg \(2018\)](#) and [Abeler et al. \(2019\)](#). [Dufwenberg and Dufwenberg \(2018\)](#) defined the reputation cost in terms of the belief the audience has about the degree of misreporting (*i.e.*, how much the decision maker has cheated), as opposed to [Abeler et al. \(2019\)](#) who defined it in terms of a mere chance of being a liar.

<sup>10</sup>See [Levitt and List \(2007\)](#) for a discussion on the use of laboratory experiments in economics.

<sup>11</sup>Given the focus of this thesis, lying games with exogenous probability of detection and sanction (*i.e.*, tax evasion game) are not discussed in detail in this review.



the cheap-talk literature (see [Crawford and Sobel, 1982](#)). It represents a situation in which one party can misrepresent the truth to the other party due to asymmetric information, like that illustrated in the previous example of a car mechanic. The setting of a Sender-Receiver Game closely resembles an economic transaction in a credence goods market (see [Dulleck and Kerschbamer, 2006](#)) where clients do not know their true needs and cannot completely verify if the experts have *correctly* identified and fulfilled those needs. In our car mechanic example, the customer may not know the underlying cause of the problem (say the car does not start either due to the fuse or the engine). If the mechanic wants to defraud this customer, he can post a high price, pretends to have replaced the engine, and simply treats the car *just enough* that it starts. From the customer's point of view, given that the car functions, he or she may not know for fact whether the service provided was in his or her best interest or not.

In the standard Sender-Receiver Game, a participant (the sender) knows the true state of the world (typically represented by two options) and it is his or her private information. The sender then sends a cheap-talk message to another participant (the receiver) about which option would yield more payoff for the receiver. The sender faces a moral dilemma because the incentives are misaligned (*i.e.*, the sender can earn more by exploiting the receiver). Finally, the decision of the receiver then determines the payoff for both parties. In this game, the sender can benefit by sending a deceptive message to the receiver regarding which option is better for him or her.<sup>12</sup> In the baseline version of this game, the fraction of senders who choose to deceive the receivers is around 40% (see [Gneezy, 2005](#); [Sutter, 2009](#)).

The second type of task represents lying behavior in a non-strategic setting. This task models a situation in which the person has to report private information without any strategic interactions (*i.e.*, a self-employed reporting income to a tax authority). A typical illustration is a die-under-a-cup paradigm *à la* [Fischbacher and Föllmi-Heusi \(2013\)](#). In a standard setting, participants are given a six-sided die and an opaque cup. Participants are then asked to privately roll the die, observe the outcome and report it to the experimenter, which determines their material payoff. Since the outcome of the die is a private information for each participant, the experimenter cannot identify liars at an individual level, but can infer about it at an aggregate level. In other words, the experimenter can know, *ex-post*, approximately how many people likely lied through statistical inference, but cannot pinpoint *exactly* who have lied. From the participant's point of view, however, he or she knows that

---

<sup>12</sup>Note the difference between lying and deceiving. Lying involves a misrepresentation of a private information. On the other hand, deception has a belief component – one can still tell the truth when one expects that the other party will believe that it is a lie. See [Sutter \(2009\)](#).

since it is only him or her who has seen the die roll, it is plausible that he or she has indeed observed the outcome he or she reports. Thus, lying is not detectable.<sup>13</sup>

One can consider two extreme predictions about how will people behave in the die-rolling task. The first considers a standard economic model of ‘*homo economicus*’ which assumes that people are rational, selfish and self-interested. Since reporting ‘5’ maximizes monetary benefit, all individuals will report ‘5’. At the other end of the spectrum, if everyone is truthful in their reports, we should expect the distribution of the reported outcomes to be uniform when we repeat the task with sufficiently large number of people.

The findings of [Fischbacher and Föllmi-Heusi \(2013\)](#) cannot be entirely supported by the standard economic model with ‘Max-U’ assumption. Assuming no one lied to their disadvantage, about 39% were honest, while 22% lied to obtain the maximum payoff (conditioned on observing a lower payoff). In addition, a number of people were *partial* liars (*i.e.*, displaying incomplete dishonesty by reporting ‘4’ instead of ‘5’). Simply put, not everyone is a utility maximizer and people are roughly categorized into three groups: the honest, the dishonest and the *partially* dishonest. This pattern of lying behavior is robust and supported by the meta-analysis of [Abeler et al. \(2019\)](#). Controlling for the financial incentives and the outcome space, [Abeler et al. \(2019\)](#) found that the average standardized report is only 0.234 (*i.e.*, 0 implies full honesty, and +1 implies full lying). These findings confirm that people are hesitant to tell a lie and thereby forgo the financial gain from being dishonest. This implies that people do experience the moral cost of lying that arises from their intrinsic preferences.

Using on the methodology of experimental economics and the tasks described, behavioral economists have identified many factors that play a role in aggravating or modulating lying behavior. In a non-strategic setting, people do not react to an increase in stake by lying more ([Abeler et al., 2019](#)), but tend to lie more when it covers financial losses (*e.g.*, [Shalvi, 2012](#); [Grolleau et al., 2016](#); [Schindler and](#)

---

<sup>13</sup>There are many adaptations of the die rolling paradigm. Some switches to a coin toss (see [Abeler et al., 2014](#)). Repeating the task for multiple rounds provides a more fine-grained indicator in terms of probability of being a liar (see [Cohn et al., 2014](#); [Cohn and Maréchal, 2017](#); [Cohn et al., 2022](#)). The Mind Game ([Jiang, 2013](#); [Kajackaite and Gneezy, 2017](#)) asks participants to first pick an outcome in their *mind*, then report if it matches with the actual outcome. This eliminates the reputation concern because there is no way for the experimenter to contradict the participant’s claim (also see [Galeotti et al., 2020](#), for a repeated mind game). The observed cheating game ([Gneezy et al., 2018](#)) follows the die rolling paradigm, except that it takes place in a computer. This allows for complete identification of cheaters at the individual level as the participants know, at the time of lying, that the experimenter will be able to directly infer about their behaviors after the experiment. Also see [Dai et al. \(2017\)](#) regarding the external validity of the die rolling task.

Pfатtheicher, 2017; Garbarino et al., 2019). As suggested by the theoretical models of perceived cheating aversion, observability has also been shown to reduce lying via reputation concerns, (*e.g.*, Gneezy et al., 2018; Abeler et al., 2019; Fries et al., 2021). By contrast, when the action cannot be observed, people become more willing to lie (*e.g.*, Jiang, 2013; Gneezy et al., 2018) and seem react to an increase in financial incentives by lying more (Kajackaite and Gneezy, 2017). There are also certain individual characteristics that correlate with the propensity to lie. Men are more likely to lie than women (Abeler et al., 2019).<sup>14</sup> Younger people lie slightly more than older cohorts (Gerlach et al., 2019).<sup>15</sup>

Other studies have examined lying in a more social setting, involving more than one person. Competition has been shown to drive dishonesty – people lie more under tournament incentives (*e.g.*, Faravelli et al., 2015; Balafoutas et al., 2020), react to prize spread by lying more (Conrads et al., 2014) or engage in unethical behaviors in a status-seeking contest (Charness et al., 2014). Lying behavior is also sensitive to the externalities it creates. People are more willing to lie when others can benefit from it (*e.g.*, Wiltermuth, 2011; Erat and Gneezy, 2012; Gino et al., 2013) and less willing to do so when it harms another participant or a third-party such as a charity (Gneezy, 2005), but as long as the stakes are low (Gneezy and Kajackaite, 2020). Groups have been shown to lie more than individuals (*e.g.*, Conrads et al., 2013; Weisel and Shalvi, 2015; Kocher et al., 2018) except when doing so creates negative externality on a third-party (Castillo et al., 2022). Relatedly, studies on peer effects showed that people tend to conform to others after being exposed to norm violations (see Keizer et al., 2008; Gino et al., 2009; Fosgaard et al., 2013; Kroher and Wolbring, 2015; Diekmann et al., 2015; Bicchieri et al., 2022), and they do so asymmetrically. That is, people tend to be more dishonest after observing norm violations but *not* more honest after observing norm compliance. For instance, Lefebvre et al. (2015) found in a lab experiment that social information about non-compliance increases tax evasion, while compliance does not reduce it. A recent study by Colzani et al. (2021) found the same asymmetry in a die rolling paradigm. Further, evidence showed that people prefer being paired with a peer of the same type to reduce the moral cost of misbehaving (*e.g.*, Gross et al., 2018; Charroin et al., 2021).

Given that there is heterogeneity in lying behavior (with and without probability of detection) and that the decision-making process in a moral domain is not a mere cost-benefit analysis (*i.e.*, it is not only a reaction to the change in audit probability and the severity of punishment as Becker (1968) suggested), behavioral economists

---

<sup>14</sup>Also see Capraro (2018) for the gender effect in the Sender-Receiver Game.

<sup>15</sup>See Bucciol and Piovesan (2011), Glätzle-Rützler and Lergetporer (2015), Maggian and Villeval (2016) and Abeler et al. (2021) for the development of lying aversion in children.

began to investigate the use of soft interventions to discourage lying. Some examples of such measures are moral nudge and reminders.<sup>16</sup> For instance, an explicit moral cue can be used to emphasize *what is the right thing to do*. Mazar et al. (2008) found that asking people to recall the Ten Commandments reduces lying. Bryan et al. (2013) found that a subtle change in a reminder from ‘Please *don’t cheat*’ to ‘Please *don’t be a cheater*’ significantly reduces cheating behavior. Along this line, Bucciol and Piovesan (2011) found a similar effect among children and documented a gender effect: a reduction caused by a request not to cheat is larger among girls compared to boys. In addition, psychologists tested the impact of an implicit moral cue, which are even more subtle than nudging or reminders. For examples, Vincent et al. (2013) and Gino and Mogilner (2014) found that participants tend to lie less when they decide in front of a mirror. Overall, these findings are consistent with the self-concept maintenance theory (Mazar et al., 2008) – people have an innate desire to be honest and want to perceive *themselves* in a positive light. Therefore, these moral cues and reminders propel an individual to behave more consistently with the moral person he or she wants to be.

Nonetheless, there are evidence reporting null effects of moral cues. Notably, Kristal et al. (2020) showed in a large-scale replication study that signing a declaration form at the top *does not* reduce lying compared to signing at the bottom. Cai et al. (2015) showed that an implicit cue of an eye does not reduce dishonesty. As for a moral nudge, Dimant et al. (2020) found no effect of a norm nudge (either using descriptive or injunctive norm) on lying behavior and argued that these moral messages do not shift norm perceptions (*i.e.*, they do not make people think that lying is more inappropriate). Contrary to this, Nieper et al. (2021) found that a norm nudge by showing good examples of others in a die rolling task promotes honesty through a shift in normative expectations. Finally, a norm nudge can also be used to *indirectly* discourage lying through a third-party enforcer. Dimant and Gesche (2021) demonstrated that an individual tends to punish a liar after being exposed to a norm nudge and showed that both descriptive and injunctive norm messages shift norm perceptions.<sup>17</sup>

The use of a moral reminder has also been tested in the field. Pruckner and Sausgruber (2013) examined the effect of legal *vs.* moral reminders on newspaper purchasing in Austria. They found that reminding people that stealing is an illegal

---

<sup>16</sup>While a nudge relies on a change in choice architecture to leverage cognitive biases to achieve an intended positive outcome without changing the set of choices (see Thaler, 2018), a reminder works through information provision. A *moral* nudge can sometimes overlap with moral reminders in the sense that people are nudged to behave morally using social or normative information.

<sup>17</sup>See Dimant and Shalvi (2022) for a discussion on the idea of ‘meta-nudging’ (*i.e.*, indirectly nudge norm enforcers to promote honesty).

act is ineffective, while simply reminding them to be honest improves civic honesty. In a tax compliance setting, [Hallsworth et al. \(2017\)](#) found a positive impact from using social norm messages to encourage payment of tax overdue in the UK and found that descriptive norm message (*i.e.*, what people do) is more effective than injunctive norm message (*i.e.*, what people think they ought to do). Similarly, [Bursztyn et al. \(2018\)](#) found a positive effect of a moral appeal in the context of debt repayment. However, there are also studies in the field that documented null effects of a moral reminder. For examples, [Blumenthal et al. \(2001\)](#) and [Fellner et al. \(2013\)](#) found no impact of a moral appeal in the context of tax compliance.<sup>18</sup>

Beyond the use of moral cues as soft intervention to reduce dishonesty, a stream of literature examined the impact of a truth-telling oath. A solemn oath is a form of commitment device that binds an agent to undertake certain behavioral acts ([Kiesler and Sakumura, 1966](#)).<sup>19</sup> Once the individual has sworn to be honest, he or she is *bound* to tell the truth when he or she will be confronted with the possibility to tell a lie. Evidence in the literature indicate that an honesty oath induces truth-telling at an individual level in a Sender-Receiver Game ([Jacquemet et al., 2019, 2021](#)) and in a non-strategic lying task ([Heinicke et al., 2019; Jacquemet et al., 2021](#)). This is because when committed to an oath, the moral cost of lying is increased which interferes with the justification process, as pointed out in [Abeler et al. \(2019\)](#).

Nonetheless, recent studies suggest that the effect of an oath may be more heterogeneous than just ‘increasing truth-telling’ or ‘reducing lies’, and that its effect may be moderated under certain conditions. [Jacquemet et al. \(2019\)](#) and [Cagala et al. \(2019\)](#) found oath does not increase truth-telling when the language is not ethically loaded. Others found that the effect of oath depends on the size and the nature of lies. [Heinicke et al. \(2019\)](#) and [Jacquemet et al. \(2021\)](#) found that oath reduces big and obvious lies but does not affect partial and small lies. On the nature of lies, [Jacquemet et al. \(2021\)](#) found in a Sender-Receiver Game that senders under oath are less likely to deceive receivers when the lie is self-serving, but are not less

---

<sup>18</sup>These null findings, particularly in tax compliance and debt repayment literature, could be that the manipulation does not work due to a lack of salience (say the moral reminder is *lost* in the letter) or it works but may not translate into behavioral changes because of, for an example, an inability to pay and/or a scarcity mindset (see discussions in [Shah et al. \(2012\)](#) and [Mani et al. \(2013\)](#)). Also see [Boonmanunt et al. \(2020\)](#) who found that reminding people of the injunctive norm (*i.e.*, the majority of people deem lying as very inappropriate) is ineffective when they are experiencing financial scarcity. See also [Antinyan and Asatryan \(2019\)](#) for a meta-analysis on the impact of nudge (*i.e.*, moral, legal, social) on tax compliance.

<sup>19</sup>An oath should be distinguished from a moral *nudge* because the latter exposes the decision maker to information about the norms of honesty without asking for any commitment. Thus, the effect of a nudge manifests itself *unconsciously* without the decision maker being aware and/or more through a moral reminder effect and a change in choice architecture.

likely to deceive when it is mutually beneficial to do so. A few studies documented a null effect of an honesty oath. Using lab and field experiments, [Koessler et al. \(2019\)](#) found no direct impact of a moral commitment on tax compliance. In a framed field experiment with villagers in Indonesia, [Prima et al. \(2020\)](#) found no effect of an oath in an asset declaration game, either on its own or when combined with verification threats.

Two notable exceptions documented a backfiring effect of an oath. [Cagala et al. \(2019\)](#) found that a neutrally-framed commitment, combined with threats of punishment, incites more cheating than not signing any commitment. This effect is driven by individuals with a high psychological reactance because they are more likely to ‘push back’ against the request. In another study, [Cagala et al. \(2021\)](#) found that committing university students to an honesty statement before an exam doubles the cheating rate than in the baseline where students do not make any commitment. The authors interpreted that signing an oath might have weakened the descriptive norm of exam-taking. In other words, the fact that an honesty statement has to be signed suggests that it is normal to cheat during exams.

In sum, the empirical evidence on the behavioral economics of dishonesty suggest that not everyone is a *homo economicus* when it comes to decision-makings in a moral domain. People have intrinsic motivations that evoke moral costs whenever their actions become misaligned with the internal rules. In addition, behavioral economists have shown that lying behavior is sensitive to the exogenous manipulation of moral costs such as an honesty oath, which is a focus in Chapter 1 and 2 of this thesis.

One of the less explored areas is the impact of an honesty oath in a setting with social interactions where lying is strategic and creates negative externality on another individual. Although [Jacquemet et al. \(2019, 2021\)](#) have explored such a setting using a Sender-Receiver Game, it was only a one-shot decision. Nonetheless, in a context such as that of a financial market, advisors often interact with investors on a repeated basis and thereby have several opportunities to defraud them by exploiting information asymmetry. This repeated interaction, besides corresponding more closely to the dynamic in the real world, can provide interesting insights into the effect of an oath where investors can learn about the honesty of their advisors while the advisors can adapt their lying strategies over time. Moreover, the existing literature have not addressed the impact of an honesty oath on the decision of a counterparty who is the victim of a lie. In a financial market setting, if the investors are aware that the advisors are under oath, would they be more willing to take their advice? How does an honesty oath work in such a market? What are the consequences of introducing an oath on advisors’ lying behavior? How does

it eventually affect the welfare of the investors? Chapter 1 fills this gap in the literature.

## **Chapter 1: Honesty Oaths in Advisor-Client Relationships: An Experiment**

The aim of Chapter 1 is to identify what makes honesty oaths work in the financial markets using a laboratory experiment. If one wonders about the relevance of oath-taking in a financial market, *bankers' oath* has been proposed to limit professional misconducts and restore public trust in the banking sector that declined in the wake of the global financial crisis of 2008 (Sapienza and Zingales, 2012). Following the Hippocratic oath for the medical profession, bankers' oath commits employees in the financial sector to the prescribed ethical standards, maintain transparency and accountability, and place a client's interest before personal ones (Boatright, 2013). This practice has been legally implemented on all employees in the banking sector in the Netherlands since 2015, followed by Belgium in 2019. As the literature have shown, such a practice may prove to be a cost-effective tool to reduce lying behavior.

In a financial market, the investor does not know if the product recommended by the advisor is indeed in his or her best interest. Being an expert with superior information, the advisor can defraud the investor by misrepresenting the true state of the world (*i.e.*, portfolio investment). On the advisor's side, he or she may consider various lying strategies – one that are *detectable ex-post* by the investor with varied probabilities, or one that remains *deniable* (*i.e.*, they can never be detected). On the investor's side, after making the decision and observing the realized outcome of the investment, he or she may be able to infer about the honesty of the advisor. This is an important aspect because, if the investor can deduce that he or she has been lied to, this can adversely affect the decision to participate in the financial market, contributing to the *lemons market* problem (Akerlof, 1970). Therefore, if an oath can maintain advisors' honesty, it may help in preventing investors from exiting the market.

More specifically, Chapter 1 experimentally investigates the impact of an honesty oath on financial advisors' lying behavior, and consequently its impact on the investors in a controlled laboratory setting. Does an oath eliminate *all* types of lies, or does it simply change the way advisors lie to their investors? Do investors trust advisors more because of the oath *per se*? Or is it the experience with their advisors that dictates investment decisions?

The main contribution of Chapter 1 is by investigating the impact of oath *beyond* reducing lies – how it influences the decision-making of the investor, who is the

victim of a lie. [Weitzel and Kirchler \(2021\)](#) conducted an audit study to examine the impact of banker’s oath on the advisors’ honesty in the Netherlands. They found that being reminded of the oath made advisors less likely to prioritize the bank’s interests (*i.e.*, pushing unnecessary loans) than in the baseline, showing that limited recall of the oath might play a role, but so might moral reminders (bankers had to explain the purpose of the oath) and a fear of punishment (customers and the institution itself might be more likely to file misconduct complaints). However, what impact the oath has on investors’ decisions still remains an open question – does being aware that the advisor is under an oath facilitate investor’s decision? Or is it the experience with the advisor that matters more for the investor? Chapter 1 explores these phenomena via a novel experimental design by varying whether the oath is common knowledge with the investors or not.

In addition, most of the existing works focused on a non-strategic individual lying task and only a few examined the impact of oath in a strategic setting where lies generate negative externality – both of which are important aspects in the financial market. While some previous studies have used a Sender-Receiver Game (*e.g.*, [Jacquemet et al., 2019, 2021](#)), it is implemented using as a one-shot game where the receiver (the victim of a lie) does not learn afterwards if the message was indeed truthful or not. On the other hand, in a financial market setting, after taking the advice of the financial advisor, the investor may discover ex-post if he or she has been lied to. This creates a learning dynamic in a repeated game setting whereby investors may start to lose trust in their advisors.

We designed a laboratory experiment using the Announcement Game developed by [Tergiman and Villeval \(2022\)](#). In this two-player game, an advisor privately observes a portfolio of investment projects, represented by three cards. Each card has a 0.5 probability to display a star, signifying a successful project. Under asymmetric information, the advisor sends a cheap-talk message about the number of stars to a player in the role of an investor. The investor then decides whether to invest in the project or not. Next, Nature draws one of the advisor’s three cards to determine the success of the investment (star or no star). In this setting, the advisor can choose whether to lie or not and what type of lie to make. A lie can be detected by the investor (*i.e.*, announcing 3 stars after observing fewer than 3) or deniable (*i.e.*, announcing 2 stars after observing fewer than 2). Thus, while detectable lies can be discovered ex-post, deniable lies cannot be uncovered *immediately* because Nature’s draw does not contradict the announcement of the advisor. We used a Stranger Matching Protocol where advisors and investors are re-matched after each period.<sup>20</sup>

---

<sup>20</sup>Although it is possible to detect a deniable liar *over time*, it is nearly impossible to uncover a deniable lie under stranger matching.



Our four between-subjects treatments varied in whether and how an oath was introduced. *No Oath* serves as a baseline condition (as described in the previous paragraph). In *Oath-Common-Knowledge*, advisors swear the honesty oath and it is common knowledge with the investors. Comparing Oath-CK with the baseline allows us to directly evaluate the impact of a common-knowledge oath. We found a drastic reduction in advisors' lies (both detectable and deniable), which consequently resulted in an increase in investment.

However, a higher investment rate can come from two factors. On the one hand, it is possible that investors anticipate advisors to announce more honestly because they are aware of the oath. In other words, the common knowledge of the oath has informational value to the investors, allowing them to trust the announcements than when they are not aware of the oath. On the other hand, because the announcements under a common-knowledge oath are predominantly honest in nature, it is possible that investors are simply responding to the experience with the honest advisors. The remaining two treatments, *Oath-No-Common-Knowledge* and *Oath-No-Common-Knowledge-Asynchronous*, disentangle the two explanations.

In Oath-NoCK, advisors swear the oath but it is not common knowledge with the investors. In Oath-NoCK-Async, all participants are the investors and face the announcements made by past advisors in Oath-CK but do not know about the oath.

The findings support the experience explanation. We found that it is the *experience of honesty* with their advisors that the investors are reacting to and not the fact that they know the advisors are under oath. In other words, investors learn about the honesty of their advisors from their experience. While the informational value of the oath plays a marginal role in determining the investors' decision, it is a crucial element for the advisors. We found that advisors lie more when oath is not common knowledge with the investors, leading to a lower level of honesty that does not maintain trust. Simply put, it is important that the advisors are reminded that the investors are aware about the oath. Our findings are particularly important given that our independent online survey conducted via Prolific with 200 residents in the Netherlands, where bankers' oath is legally binding, indicates that 68% were not aware that Dutch bankers were required to take an oath. Thus, the findings of Chapter 1 have implications regarding the implementation of oath or professional codes of conduct: it must be accompanied with a consistent and well-planned public-relations strategy to ensure awareness both in and out of the organization.

In all, Chapter 1 explored the impact of an honesty oath on lying behavior in a strategic setting with social interactions and negative externality – one that resembles a financial market. Our novel contribution is by varying the common knowledge element of an oath to examine its impact on lying behavior and consequently invest-

ment decisions. We identified that even though oath in itself has small informational value for the investors, the common knowledge element is a necessary condition in sustaining advisors' honesty in order to create an experience that maintains trust with the investors.

Another unexplored area in the literature of oath-taking is the impact of an honesty oath when individuals are in a group setting. Given that an oath is being adopted by practitioners to reduce professional misconduct in an organization, it is important to examine if such an intervention is effective in a *collective setting* where individuals can communicate with each other and may need to jointly make ethical decisions. Chapter 2 is the first to investigate this phenomenon in a laboratory experiment.

## Chapter 2: Individual Oath-Swearing and Lying in Group

Continuing to explore the impact of an honesty oath from another standpoint, Chapter 2 aims to test whether an oath reduces lying when individuals are in a group setting. This is a relevant context to investigate for two reasons. First, financial frauds and scandals often involve *group of individuals*, not a *single person*. Enrons, Madoff Ponzi scheme and Wells Fargo are some examples where groups of individuals are involved in the intricate webs of financial crimes. Second, after an organization has implemented an oath or a professional code of conduct on their employees, they will be in a working environment where they may face the same objective and make ethical decisions with others within their organization. Would an honesty oath effectively discourage dishonest behavior in such a group context?

The literature on lying behavior suggests that groups tend to be more dishonest than individuals (*e.g.*, Weisel and Shalvi, 2015; Kocher et al., 2018). Several reasons explain the *dishonesty shift* in groups. First, studies on group decision-making have found that groups are more *sophisticated* than individuals and are more likely to make a self-interested choice in an economic game (see a review by Charness and Sutter, 2012). For instance, groups choose a lower number in a beauty contest game (Kocher and Sutter, 2005), send and return less money in a trust game (*e.g.*, Cox, 2002; Kugler et al., 2007), make and accept smaller offer in a ultimatum game (Bornstein and Yaniv, 1998). Related to lying, Sutter (2009) found in a Sender-Receiver Game that groups are more likely to use a sophisticated deception than individuals (*i.e.*, telling the truth believing that the message will not be followed). Second, the group setting allows for *diffusion of responsibility* (*e.g.*, Falk and Szech, 2013; Falk et al., 2020). This occurs when the action by an individual towards a group decision becomes less pivotal, leading to a lower perceived individual accountability. Thus, individuals exploit a moral wiggle room (Dana et al., 2007) in favor of their immoral

action by hiding behind others (*e.g.*, [Conrads et al., 2013](#); [Weisel and Shalvi, 2015](#); [Bauer et al., 2021](#)). Third, groups may lie more than individuals because *others can benefit* from it (*see e.g.*, [Wiltermuth, 2011](#); [Gino et al., 2013](#)). However, a recent study from [Kocher et al. \(2018\)](#) found that groups lie to the same extent irrespective of whether others can benefit from it or not. Lastly, previous studies suggest that individuals in a group tend to lie more because the setting allows them to learn about the empirical norm of honesty (*i.e.*, what others do) by observing the action of their peers ([Bicchieri et al., 2022](#)) or communicating with them ([Kocher et al., 2018](#)). That is, the dishonest shift in groups results from the *erosion of the norm of honesty*. Chapter 2 focuses on the last two explanations: how an oath would work when individuals can communicate with others in their group, and how it interact with the incentive structure.

Chapter 2 makes a twofold contribution. Given that previous literature have shown that an honesty oath is effective in reducing lying behavior at an individual level, this chapter is the first to examine whether this impact *persists* when individuals are in a group. On the one hand, oath increases the moral cost of lying and makes the honesty norm more salient. An individual may thus promote honest behavior within their group. On the other hand, because group members can communicate, they may exchange justifications and convince each other to break the oath. This is possible since the literature on peer conformity showed that individuals tend to lie out of conformity after observing the norm violations of their peers (*see e.g.*, [Gino et al., 2009](#); [Fosgaard et al., 2013](#); [Innes and Mitra, 2013](#); [Kroher and Wolbring, 2015](#)). Thus, this chapter contributes in relation to the effect of peer conformity and lying under oath in a group setting. In addition, this chapter investigates whether an oath intervention interacts with the incentive structure faced by the group members. This is relevant for organizations that impose an oath on their employees because group incentives may interact with the intervention in such a way that its impact may be crowded out. Chapter 2 provides a novel evidence that the impact of an honesty oath against lying in a group can be crowded out by an incentive structure that imposes *social pressure* on individuals to conform to the group when they can influence each other via group communication. This cautions practitioners regarding the interactions between an oath intervention and the economic incentives for their employees.

Following a mixed design of [Kocher et al. \(2018\)](#), the experiment used an observed cheating game of [Gneezy et al. \(2018\)](#). Participants observed the video of the outcome of a die. They were asked to memorize the outcome and report it later to earn the payoff. The three parts (within-subjects) varied in social interactions. In part 1, participants played the game individually without any social interactions.

In part 2, participants were randomly matched to form a group of three. Group members observed the same video and could anonymously communicate for five minutes via computer, after which they reported the die outcome *individually* as in part 1. Finally, in part 3, participants played the game in an individual setting as in part 1. The between-subjects dimension was applied at the beginning of part 2. The four treatments varied in oath intervention (with or without oath) and the incentive structure for the group members (with or without payoff commonality). The honesty oath committed participants to be honest and to always tell the truth for the remaining of the experiment at the beginning of part 2 (before receiving the instructions). The payoff commonality rule required all group members to report the same number to earn the payoff. Otherwise, they earned nothing. Therefore, this yields four between-subjects treatments: BaseNoPC, BasePC, OathNoPC and OathPC.

The main finding of Chapter 2 is that individual oath-swearing reduces lying in group, *but* the magnitude of the impact depends on the incentive structure. Oath causes a larger reduction in lying when group members do not earn a common payoff than when they do. The fraction of liars is higher in BaseNoPC (91.4%) compared to OathNoPC (76.5%). In fact, the impact of oath under the payoff commonality rule is marginal (94.9% in BasePC vs. 86.2% in OathPC). In addition, the fraction of liars in part 2 is higher than in part 1 for all treatments, with the difference being the least in OathNoPC. These findings are in line with the notion that the payoff commonality rule can exert pressure on individuals to conform to the group, crowding out the impact of the oath on honest reporting.

The analysis of lying behaviors in part 2 and 3 shed light on the reasons behind the larger impact of oath under no payoff commonality. The rationale is that if individuals followed their group members and lied out of pure conformity in part 2, one should expect them to continue to conform when they were alone in part 3. On the other hand, if some individuals also acted out of group pressure in part 2, one expect less lying when these individuals were alone in part 3 because the action in the group setting was forced. While the pure conformity exists in both incentive structures, group pressure should exist only under payoff commonality.

The data indicate that the effect of pure conformity is very strong in the baseline conditions. Lying rates in an individual setting after group interactions are similar across BaseNoPC and BasePC (91% and 92%, respectively). More importantly, these lying rates do not differ from the rate in a group setting of their own treatment. This implies that pure conformity drives lying in a group setting in the baseline conditions, while the group pressure under payoff commonality may play a minor role without an oath. However, there seems to indeed be group pressure under

the payoff commonality rule *with* an oath. Controlling for pure conformity (as proxied by the number of liars in the group setting in part 2), individuals in the OathPC were about 12 percentage points less likely to lie later when they were alone, compared to all the other treatments. This implies that a high lying rate under oath in a group setting with payoff commonality is driven by pure conformity *and* group pressure, which crowds out the impact of the oath. While a concern for efficiency cannot be ruled out completely, it should co-exist with a peer pressure to lie. Overall, the findings of Chapter 2 highlight that an oath intervention can interact with the incentive structure governing the group. Therefore, to be effective in an organizational context, firms must implement an oath intervention in close connection with the incentive schemes.

So far, Chapter 1 and 2 showed that an honesty oath, as an intervention to raise a moral cost of lying (Abeler et al., 2019), can discourage lying behaviors in a strategic setting where lying creates negative externality and when individuals are in a collective setting, respectively. They contributed by identifying when an oath works and, more importantly, when its effect can be moderated. The findings of Chapter 1 and 2 provide important practical implications concerning the implementation of an oath in a market setting and within an organization respectively in order to maximize its desirable effect against unethical behaviors.

The findings of Chapter 1 and 2 are consistent with the theoretical frameworks of perceived cheating aversion (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019) which suggest that individuals suffer a moral cost of lying from intrinsic preferences and reputation concerns and they incorporate this intrinsic cost into the decision to lie. Chapter 3 takes a different perspective from the two initial chapters by focusing on what happen *after* people have lied. More precisely, after people have incurred a moral cost of lying, does its role in the decision-making process end there? Or does it continue to influence how people use the money they have earned *dishonestly*? If the influence of a moral cost persists in the way individuals make their economic decisions, do they use *unethical money* differently from that earned ethically? Do they treat it more like a windfall gain or that earned from costly effort? Chapter 3 investigated these questions in the context of an economic decision – how individuals take risk with their money.

### **Chapter 3: Risk-Taking with Unethically Earned Money**

Chapter 3 aims to understand how people use the money they have earned through unethical means. More specifically, it investigates whether unethically earned money is treated more as a windfall gain or as the result of costly effort in the context of risk-taking. This is very well-suited to study using experimental methods because it

is nearly impossible to observe in the real world how individuals use their dishonestly earned money and if they use it differently from money they earned through ethical sources. This reflection relates to mental accounting (Thaler, 1985, 1999; Abeler and Marklein, 2017) which shows that the sources of earnings influence how individuals spend money. This type of behavior violates the principle of economic fungibility (Abeler and Marklein, 2017), which posits that any unit of money is substitutable (*i.e.*, a dollar from one account is interchangeable with the same from another account). If people take risk with their money differently depending on whether it was earned ethically or not, it may have implications on risk taking in domains such as financial markets where the exploitation of asymmetric information may be tempting.

A recent study by Imas et al. (2020) explored a violation of fungibility of unethical money in the context of charitable giving. In several experiments, they found that dishonest individuals engaged in motivated mental money laundering – a simple exchange of physical bills of unethically earned money with *clean money* significantly reduces the tendency with which dishonest individuals donate their money to a charity. Chapter 3 differs from Imas et al. (2020) in two aspects. First, it investigates a violation of the fungibility principle in the context of risk taking, which is unrelated to the moral domain. Second, Imas et al. (2020) studied a motivated reasoning engaged by dishonest individuals to dissociate unethical money from its source. In contrast, by studying how much risk individuals are willing to take with dishonestly *vs.* honestly earned money, Chapter 3 explores a *non-motivated* violation based on mechanisms identified when studying violations of fungibility between money earned through luck or effort.

The literature in behavioral economics have shown that people tend to use money earned through costly effort differently from a windfall gain (*i.e.*, Hoffman et al., 1994; Cherry et al., 2002; List and Cherry, 2008; Hvide et al., 2019), which has been explained by a sense of entitlement because incurring costly effort induces a feeling of ownership. For risk-taking behavior, a sense of entitlement should induce risk averse behavior. In contrast, obtaining a windfall gain imparts a weak sense of entitlement, leading to a higher tendency to take risk. As to how one takes risk with unethically earned money is not very straightforward for two reasons.

On the one hand, people may perceive the money earned through dishonest means or lying as something easily earned like a ‘*house money*’ (Thaler and Johnson, 1990) because it is *objectively* not their money (*i.e.*, it belongs to other people). If this is so, one can expect a high risk taking with unethically earned money because people may treat it more like a windfall gain. On the other hand, the individual may incur a moral cost to obtain unethical money due to intrinsic lying aversion or

reputation concerns associated with perceived cheating aversion (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019). Such moral costs might act like an effort cost which, then, induces risk aversion. In other words, dishonest individuals may justify a sense of entitlement to the unethical gain and thus, behave *as if* they were actually entitled to it. This reasoning is based on the psychological cost hypothesis of Thielmann and Hilbig (2019) who argued that although dishonest individuals are not *objectively* entitled to the unethical gain, they may justify *subjectively* that it is ‘*theirs*’ because they have incurred a moral cost to obtain it. If this is so, suffering a moral cost to obtain money may reduce the willingness to put it at risk in future decisions.

We conducted an online experiment where participants first earned an endowment, after which they made a risk-taking decision that could reduce the earnings realized in the first stage (in the event of loss). Our three between-subjects design varied in how participants earned the endowment (either through luck in a binary lottery, through costly effort in a real-effort task, or unethically from a lying task). We found that in the context of risk-taking, individuals treat dishonestly earned money more like a windfall gain from luck than as an effort-based gain. The effect is especially prevalent among risk averse liars.

Nonetheless, this finding could result either by design from a selection effect of low-moral cost people (because lying was observable *ex-post* by the experimenter) or the fact that a moral cost does not generate any sense of entitlement. To disentangle between these two possibilities, we conducted a follow-up experiment, keeping the same selection effect, in which we increased the moral cost (*ex-post* to the lying) using social norm responses elicited from a separate sample in the same subject pool. In particular, participants in the follow-up experiment were exposed to the injunctive norms of the lying task – that the truthful (lying) option was deemed as socially appropriate (inappropriate). This moral cost manipulation was done after the lying task but before the risk task. We found that increasing the moral cost eliminates the difference between risk-taking with dishonestly earned money and with money earned from costly effort. This implies that moral costs induce an entitlement effect. The findings provide implications concerning the role of intrinsic cost of lying to limit the reckless use of money, such as gambling for an example. This can be relevant in a context such as a financial market where fraud is not uncommon. Egan et al. (2018) documented that 44% of advisors who were fired for misconducts found a new job in the same year. Moreover, there was a matching market for misconducts – advisors with misconduct records sought out firms that hired people like them, while these firms were less likely to fire advisors for new misconducts. If the moral cost of obtaining unethical money is low, it then implies

to a tendency to take risk, which can indirectly influence the way advisors allocate portfolio for their clients (see [Foerster et al., 2017](#); [Kling et al., 2019](#); [Linnainmaa et al., 2021](#)).

\*\*\*

To conclude, this thesis contributes to the literature of behavioral economics of dishonesty by investigating the influence of a moral cost, arising from individual intrinsic motivations, on the decision to lie and whether such a cost persists to influence the decision-making process after one has lied. While we still need to be cautious about extrapolating the findings from laboratory and online experiments in a more complex environment, this thesis advances our understanding of when and why people lie, how can we optimally design interventions to reduce dishonesty at an individual and the collective setting, and what is the role of a moral cost on the economic use of money in a risk domain.

In all, the findings presented in this thesis provide a reflection of the moral decision-making process of a *homo moralis*, which resonates with the moral philosophies of Adam Smith and Immanuel Kant. As social creatures, we behave in the manner that does not always maximize our own material benefit. We care not only about the welfare of others but also how our actions may be misaligned with the normative rules and our innate moral motivations. Our intrinsic motivations *and* the shared beliefs we hold with others construct the moral compass that provides us with guidance for socially fit and proper behaviors, enabling us to live with each other in harmony. Nevertheless, as Smith pointed out about the humanly nature of *self-love*, our passions may become our own demise by which we are purblind and act in discord with the moral rules and duties. Are we able to restrain our selfish desires and promote a human socioeconomic betterment? Or are we lost because we ignore our moral compass?





# Chapter 1

## Honesty Oaths in Advisor-Client Relationships: An Experiment<sup>1</sup>

### 1.1 Introduction

A number of large and well publicized frauds in the financial sector (for example the Madoff, Enron, and Wells Fargo scandals) have led to a decrease in the public's trust towards the financial sector (Guiso et al., 2008; Sapienza and Zingales, 2012). In fact, “trust issues” are endemic to financial markets, leading individuals to exit these markets or not participate in them altogether, forgoing possibly lucrative opportunities. For instance, Johnson et al. (2015) found that a reason behind a sluggish mortgage re-financing was a feeling of suspicion- borrowers perceived the offer “too good to be true”, anticipated hidden costs and therefore simply left money on the table. Moreover, Johnson et al. (2015) found that incentive programs designed to encourage the offer uptake backfired, which suggests that distrust can be hard to overcome. To this end, a banker's oath has been proposed as a tool to reduce fraud and restore public trust in the financial sector. Closely following the Hippocratic Oath for the medical profession, a banker's oath aims to commit employees in financial institutions to prescribed ethical standards, maintain transparency and accountability, and place a client's interest before personal ones (Boatright, 2013). The Netherlands was the first country to implement such a practice, and since 2015 all employees in the Dutch banking sector are legally obliged to swear this oath prior to starting their jobs.<sup>2</sup> A similar intervention was adopted in Belgium in 2019, but still remains an uncommon practice in the other countries.

---

<sup>1</sup>This Chapter is joint with Chloe Tergiman and Marie Claire Villeval.

<sup>2</sup>Loonen and Rutgers (2017) provides a summary on the introduction of bankers' oath in the Netherlands. See also: <https://www.tuchtrechtbanken.nl/en/about-the-bankers-oath/>

An immediate question follows: what is the impact of such an oath, both on the individuals who take the oath, and on investors who rely on those individuals' honesty to make investment decisions? This chapter aims to answer these questions using a laboratory experiment and an online survey of residents of the Netherlands to show the empirical relevance of our findings.

We designed a laboratory experiment using the Announcement Game developed in [Tergiman and Villeval \(2022\)](#). It differs from a “standard” Sender-Receiver Game in a number of dimensions that we argue are important in trying to understand the impact of oaths in financial markets specifically. In the Announcement game, players in the role of a sender endogenously choose what kinds of lies to make. These lies can be detectable (with various levels of detectability) or deniable. In addition, players in the role of a receiver can, in certain circumstances, learn whether a lie has occurred but not what the true state of the world was. These features are relevant in studying the impact of oaths in financial markets where the types of lies that can be made are richer than what standard Sender-Receiver Game can allow for.

In the Announcement Game, a financial advisor is paired with an investor. Financial advisors receive three cards which represent a portfolio of investment projects. Each card has a 0.5 probability to display a star, signifying a successful project. Under asymmetric information, the advisor, who can see the face of all three cards, sends a cheap-talk message about the number of stars to a player in the role of an investor. The investor then decides whether to invest in the project or not. Next, Nature draws one of the advisor's three cards to determine the success of the investment: the investment is a success if the drawn card has a star on it, and is a failure if the drawn card is blank. The novel feature of the Announcement Game is that the advisor can endogenously choose what kinds of lies to make. For example, lies can be *detectable*, which happens when they announce 3 stars after observing fewer than 3 – in these situations the advisor might get caught if Nature draws a blank card. But lies can also be *deniable*, for example announcing 2 after observing fewer than 2 stars. In these situations, advisors can maintain plausible deniability regardless of Nature's draw since whatever the draw, it will not contradict the 2 stars announcements. One of the main results of [Tergiman and Villeval \(2022\)](#) is that deniable lies are widespread, even after the introduction of reputation, showing that traditional market mechanisms may not be strong enough to root out lies and misrepresentations, particularly for deniable lies and lies that have a lower probability of detection. The Announcement Game is uniquely suited to allow us to examine how an oath impacts various types of lies in a strategic setting with asymmetric information and negative externalities, which we believe captures important elements of financial markets.

Our four between-subjects treatments vary whether and how an oath is introduced. The *No Oath* treatment serves as a baseline in which advisors do not swear any oath. In the *Oath Common Knowledge (Oath-CK)* treatment, advisors swear an oath, and this is common knowledge among all participants. Comparing data from the No Oath and Oath Common Knowledge treatments allows us to directly measure the impact of such an oath on advisors' and investors' behavior. The data show that common knowledge oath-taking has a dramatic impact on all types of lies. The fraction of advisors who make detectable lies drops significantly from 71% to 12%.<sup>3</sup> More surprising is the impact on the fraction of advisors who engage in deniable lies. The fractions go from 87% to 23% after the common-knowledge oath is introduced.<sup>4</sup> Investors invest significantly more under the Oath-CK treatment, and their behavior shows that they interpret 2 and 3 stars announcements as more credible than when there is no oath (investments after 2 stars announcements go from 71% to 87% after the oath is introduced, and investments after 3 stars announcements go from 78% to 98%). In short, compared with the baseline, in the Oath-CK treatment, announcements are more truthful, and investors treat announcements as being more truthful.

While the common-knowledge oath has a dramatic impact on behavior, the question of why the oath works remains open. From the investors' standpoint, are investors anticipating more honest behavior from advisors who take an oath and therefore treat announcements as more honest? Or are investors reacting to the honesty of announcements independently of the oath? That is, are they reacting to their experience? From the advisors' standpoint, is the oath *per se* leading advisors to be more honest, or does the common-knowledge element play a role and if so to what extent?

To explore the answers to these questions, we conduct two treatments in which we remove the common knowledge aspect of the oath. In the *Oath No Common Knowledge Asynchronous (Oath-NoCK-Async)* treatment, investors are paired with the past advisors from the Oath-CK treatment described above. All subjects in these sessions are investors who are given the instructions from the Oath-CK treatment, except for the portions regarding the oath. Thus, they face announcements from past advisors who had taken a common-knowledge oath. Studying the behavior of investors in the Oath-NoCK-Async and comparing it to that of

---

<sup>3</sup>As a comparison point, in [Tergiman and Villeval \(2022\)](#), the introduction of reputation leads to a much lower drop in detectable lies: from 82.9% without reputation to 50% with reputation. This shows that the oath has a more powerful impact on detectable lies.

<sup>4</sup>In [Tergiman and Villeval \(2022\)](#), the introduction of reputation has no impact on deniable lies. If anything, reputation pushes advisors to engage in deniable lies strategy as the fraction of deniable liars increases from 85.7% without reputation to 97.4% with reputation.

investors in the Oath-CK treatment, we show that, largely speaking, it is not the oath *per se* that investors are reacting to in the Oath CK treatment, but rather to their experience in terms of the honesty of announcements: honest announcements bring forth trust in these announcements even in a setting where lies, in particular deniable ones, can be very difficult to detect. Thus, the common-knowledge element of the oath is not material to investors' behavior.

Finally, in the *Oath No Common Knowledge (Oath-NoCK)* treatment, advisors swear an oath, but investors are kept in the dark regarding the existence of the oath, a fact that the advisors are aware of. Both advisors and investors then play the Announcement game. Comparing the Oath-NoCK treatment with the Oath-CK treatment allows us to show that common knowledge impacts advisors' behavior, and therefore investors' behavior: when the oath is not common knowledge, advisors lie more, and investors invest less. That is, a no-common-knowledge oath is effective only on a subset of advisors.

We then conducted an online survey via Prolific with 200 respondents residing in the Netherlands, where employees of the Dutch banking sector have to take an honesty oath. We evaluate the extent of public awareness about the oath because it might dictate the impact of the oath itself. Among respondents who reported to have a financial advisor, 68% were not aware that Dutch bankers were required to take an oath. This means that even in a country where such an oath is a legal obligation, the awareness is not universal, which could lessen the desirable impact of the oath.

Our laboratory and survey findings provide important implications for financial institutions concerning the implementation of bankers' oath. Beyond showing the impact of oath intervention on lying behavior, our experimental design demonstrates that a common knowledge about the oath plays a crucial role in determining its impact. Because common-knowledge is a factor that impacts how truthful an advisor is for a non-negligible proportion of advisors, to achieve an optimal outcome, an oath intervention must be accompanied with a consistent and well-planned public-relations strategy to ensure awareness both inside and outside the organizations, which eventually results in an experience of honesty for the clients. In short, an oath intervention with common knowledge brings forth honesty and trust, both of which are the essential foundations of well-functioning markets.

The remainder of this chapter is organized as follows. Section 1.2 briefly reviews the related literature. Section 1.3 outline the experimental design and procedures. Section 1.4 describes the behavioral conjectures. Section 1.5 reports the results of the experiment. Section 1.6 discusses these findings and concludes.

## 1.2 Related Literature

This chapter makes several important contributions to the literature. The literature on oaths in the financial sector is remarkably thin, and our literature search identified only [Weitzel and Kirchler \(2021\)](#) as studying the impact of oaths on bankers' honesty. Understanding whether honesty oaths have an impact in the field among bankers is of paramount importance if one is to improve the honesty of a sector that suffers from a lack of trust from the public. The authors show exactly that by conducting a large-scale audit study with Dutch banks. They found that being reminded of the oath makes financial advisors less likely to prioritize the bank's interests (*i.e.*, pushing unnecessary loans) than in the baseline, showing that limited recall of the oath may play a role, but so might moral reminders (bankers had to explain the purpose of the oath) and a fear of punishment (customers and the institution itself may be more likely to file misconduct complaints). However, what impact the oath has on investors' decisions remains an open question, as does the role of the common knowledge aspect of the oath *per se*. This chapter fills that gap.

In addition, on the advisors' side, this chapter relates to the literature on biased advice. Evidence in the field show that financial advisors often give biased recommendations to their clients because the incentive scheme often rewards them based on the sales of certain types of financial products (see [Mullainathan et al., 2012](#); [Anagol et al., 2012](#); [Pool et al., 2016](#); [Foerster et al., 2017](#)).<sup>5</sup> As a consequence, this gives rise to self-serving tendencies where advice become biased and may not be in the best interest of the clients. Thus, this chapter contributes to this stream of literature by showing whether and under what conditions an honesty oath can reduce the tendency that an expert defrauds naive clients, which could thereby de-bias the advice and improves clients' welfare as a consequence.

We are also related to studies on the impact of oaths in other contexts, which have shown that an honesty oath can be a very effective (not to mention cost-effective) tool to reduce lying. These experimental designs either use an inference-based non-strategic lying task (see *e.g.*, [Heinicke et al., 2019](#); [Jacquemet et al., 2020, 2021](#)), or a Sender-Receiver Game (*e.g.*, [Jacquemet et al., 2019, 2021](#)). The more relevant ones to our present work consist of those in which the receiver is hurt by the lies. [Jacquemet et al. \(2021\)](#) showed that an oath increases truth-telling in the Sender-Receiver Game when lying harms the receivers. However, when lying is mutually beneficial, there is no impact of the oath. There are notable exceptions to these findings. For example, [Koessler et al. \(2019\)](#) found no impact of oath taking on tax compliance, and [Prima et al. \(2020\)](#) showed that in an asset-declaration game in a

---

<sup>5</sup>Also see [Cain et al. \(2005\)](#) and [Kling et al. \(2019\)](#) for biased advice in the lab.

framed field experiment with villagers in Indonesia, an honesty pledge does not work on its own or when combined with verification threats. Further, the impact of an oath may vary depending on the language and frame that is used. [Jacquemet et al. \(2019\)](#) found that oath is ineffective when the instruction is not loaded (*i.e.*, not explicitly calling a lie ‘a lie’), while truth-telling increases under a loaded instruction. In a similar vein, [Cagala et al. \(2019\)](#) found that a neutrally worded commitment statement (*i.e.*, declared not to violate the rule in the instruction) does not reduce lying, while the one that prescribes ethical standards does. Besides, oath has been found to backfire (see [Cagala et al., 2019, 2021](#)).

We contribute to the above literature in terms of the choice of game. We use the Announcement Game from [Tergiman and Villeval \(2022\)](#) which is a strategic lying game that differs from the standard Sender-Receiver Game in various dimensions. Notably, senders in the Announcement Game can vary the detectability of their lies, depending on the state of the world which is a private information. On the other hand, the receiver may or may not learn afterwards whether he or she has been lied to. These two features are novel to the Announcement Game and are well-suited to study the effect of oath in a financial market setting. In addition, the repeated-game setting of the Announcement Game allows us to explore whether the effect of oath is durable. Previous studies examined the impact of oath using a one-shot game, which yields insights only on its immediate effect. Only [Peer and Feldman \(2021\)](#) showed that the effect of oath on lying is stable over time in the multiple round non-strategic lying task. Thus, we also contribute to this investigation using a strategic lying game where players interact repeatedly and lying generates a negative externality.

Lastly, beyond the choice of game *per se*, the novel feature of our experimental design varies whether the oath is common knowledge with the victims of lies, thereby allowing us to ask new questions and evaluate what drives behavior under oath, both for investors and for advisors. Common knowledge may further encourage advisors to behave more honestly when they know that the investors are aware that they are under oath. Meanwhile, this information may allow investors to put more trust in their advisors. Either way, if the effect of oath is further reinforced by public awareness, this would have practical implications regarding the publicity of oath intervention in relevant institutions.

## 1.3 Experimental Design and Procedures

### Design

To study the impact of oath on truth-telling and the way people lie in market, we use the Announcement Game of [Tergiman and Villeval \(2022\)](#). In this game, participants are either assigned the role of an advisor or an investor. Roles are fixed for the whole session. The game was played under a finite but unknown horizon to avoid end game effects. In particular, the number of periods (18) was determined before the experiment. Participants were informed that they would play a minimum of 10 periods and a maximum of 30 periods, and this was common knowledge.

In each period, advisors and investors start with an endowment of 30 and 100 tokens, respectively. At the beginning of the period, a random draw determines a set of three cards for each advisor, each of which has an independent 0.5 probability to display a star, which indicates a successful project. The number of stars (0, 1, 2 or 3) is private information to the advisor. After observing the number of stars, the advisor sends a cheap-talk message to the investor regarding his or her number of stars and thus, can misreport the actual number. The investor then decides whether to invest or not his or her endowment with the advisor. Next, Nature selects one of advisor's three cards to determine whether the investment is a success (if the drawn card displays a star) or a failure (if it is blank).

Irrespective of the investor's decision, the advisor and the investor learn about the outcome of Nature's draw. The payoff of the advisor is state independent; the advisor earns 230 tokens (the initial 30 tokens plus 200 tokens) if the investor invests, otherwise he or she earns 30 tokens. The payoff of the investor depends on the decision to invest, and the outcome of Nature's draw. If the investor decided not to invest, he or she keeps the initial endowment of 100 tokens. If he or she decided to invest and a star is drawn, the investment is successful and he or she earns 300 tokens ( $100-100+300$ ). And if he or she decided to invest and a blank card is drawn, the investment fails and he or she earns 30 tokens ( $100-100+30$ ). We use a Stranger Matching Protocol- the advisor and the investor are randomly re-matched to form a new pair after each period.<sup>6</sup> At the end of the session, the program randomly selected one period to count for an additional payment for the Announcement Game.

As in [Tergiman and Villeval \(2022\)](#), to facilitate their understanding, before the Announcement Game, participants played eight practice periods of the Truthful An-

---

<sup>6</sup>The reason for this choice is that [Tergiman and Villeval \(2022\)](#) showed that this configuration is the one that suffers the most from dishonest behavior.



nouncement Game in the role of an investor, paired with a truth-telling computer.<sup>7</sup> These practice periods were aimed to help participants encounter all possible cases of the announcements corresponding to the probabilities specified in the Announcement Game. At the end of the practice rounds, participants made five incentivized decisions. In period 9, participants decided whether or not to invest in each of the four scenarios (*i.e.*, 0, 1, 2, 3 stars). In period 10, they made an investment decision not being informed of the number of stars. Investment decisions in period 9 and 10 allow us to establish a benchmark of risk attitudes to study investors' behavior in the Announcement Game. At the end of the session, one of the two periods was randomly selected to count for the additional payment for the Truthful Announcement Game. Instructions are available in Section 1.A of the Appendix. Below we describe our four between-subjects treatments.

**No Oath Treatment:** In the No Oath treatment, participants in the role of the advisor did not take any oath at the beginning of part 2.<sup>8</sup>

**Oath Common Knowledge (Oath-CK) Treatment:** In the Oath-CK treatment, participants in the role of the advisor were required to swear an oath. The instructions for all participants stated that those who would be assigned the role of advisor would be asked to take an oath by which they swore upon their honor that during the experiment they would be honest and always tell the truth. The oath-taking stage took place after participants discovered their role in the Announcement Game. Advisors swore an oath by retyping the text of the oath on their computers. During this stage, those in the role of the investor were informed on the computer screen that participants assigned to the role of the advisor were taking an oath. After the oath-taking stage, the investors were informed that all the advisors had sworn the oath. The taking of the oath was thus common knowledge.

Our oath-taking procedure differs from how it has been typically implemented in the literature (see [Jacquemet et al., 2019](#)) in three ways. First, we made oath-taking compulsory for all advisors while most previous studies made it voluntary. This is a design choice because we believe that in reality individuals do not have the option of not signing, had they have to sign one (like in the case of Dutch bankers). In addition, when oath was voluntary, the decline rate was negligible and almost everyone signed the oath. Second, the oath-taking stage was implemented

---

<sup>7</sup>A difference in the Truthful Announcement Game with [Tergiman and Villeval \(2022\)](#) is in the number of periods: 16 practice periods vs. eight in this study.

<sup>8</sup>This treatment corresponds to the No Reputation treatment in [Tergiman and Villeval \(2022\)](#). The only difference is in the number of periods (27 vs. 18 in this study).

in the middle of an experimental session, and not upon arrival at the laboratory. In addition, our participants have already read the instructions and thus knew the content of the Announcement Game, as opposed to the literature where subjects swore an oath without knowing the content of the experiment. We believe that this procedure makes the oath salient in the Announcement Game. Implementing it too early will not make it salient enough since our participants would first play the Truthful Announcement Game, which is irrelevant to the oath. Lastly, the oath was not taken in private, compared to previous laboratory studies that implemented it in a sealed room with an experimenter. This design choice allows us to vary the common knowledge aspect of the oath with the investors who are potential victims of the lies.

**Oath No Common Knowledge Asynchronous (Oath-NoCK-Async) Treatment:** All participants in the Oath-NoCK-Async treatment were assigned the role of the investor. When came the time for part 2 instructions, participants were told that they were about to be given an excerpt of the instructions that were given to participants in the past sessions. Those instructions were those from the Oath-CK treatment, and were identical to them except for the reference to the honesty oath.<sup>9</sup> After reading those instructions, participants were told they were going to be matched with the advisors from those past sessions and would face their announcements and then Nature's draw in those sessions.<sup>10</sup> Investment decision affects only the earnings of the investors and not the past advisors in the Oath-CK treatment. To ensure that participants had enough opportunity to learn from experience, we extended the number of periods in this treatment to 27. This was unknown to the participants.

**Oath No Common Knowledge (Oath-NoCK) Treatment:** In the Oath-NoCK treatment, the instructions handed out to all participants were identical to those of the No Oath treatment. However, after discovering their roles in the Announcement

---

<sup>9</sup>As investors in the Oath-CK treatment knew about the oath *and* experienced higher level of honesty with their advisors, we chose to maintain the same empirical distribution of the announcements made by the advisors in the Oath-CK treatment. The instruction of the Oath-NoCK-Async treatment did not use any forms of deception.

<sup>10</sup>For each period, we randomly selected one period among all the decisions made by the past 77 advisors in the Oath-CK treatment (which showed no significant time trends in advisors' or investors' behavior) and showed the investors that announcement. After the investment decision was made in the Oath-NoCK-Async treatment, we showed them Nature's draw from the past game. Thus, in the Oath-NoCK-Async treatment, we preserved the same empirical distribution of the announcements made by the advisors and Nature's draws in the Oath-CK treatment.

Game, all the participants in the role of advisors were asked to take the honesty oath, while those in the role of the investor were asked to type a neutral sentence during the oath-taking stage.<sup>11</sup> Advisors were made aware that all advisors were asked to take the oath, and were also made aware of the asymmetry in information with respect to the investors.

After completing the Announcement Game, all participants answered incentivized questions about their beliefs depending on their role. Investors were asked to estimate the proportion of honest announcement for each number of stars made to them during the experiment (*i.e.*, the first order belief about the honesty of the advisors). Advisors were asked to estimate the guesses made by the investors about the proportion of honest announcement for each number of stars (*i.e.*, the second order belief of the advisors) and the proportion of honest announcement for each number of stars made by all advisors in the session (*i.e.*, the first order belief about the advisors). They could earn an additional payoff of 0.50 Euros if their answer in the randomly selected guesses was within an interval of 5% point. Finally, participants received feedback about their payoffs for each part, the randomly selected belief question and answered a standard socio-demographic questionnaire.

## Procedures

The experiment was run in-person at GATE-Lab in Lyon, France between September and December 2021. We ran in total 26 sessions: 7 sessions with 152 participants in the No Oath treatment, 7 sessions with 154 participants in the Oath-CK treatment, 5 sessions with 77 participants in the Oath-NoCK-Async treatment<sup>12</sup>, and 7 sessions with 146 participants in the Oath-NoCK treatment.<sup>13</sup> All 529 participants (50.7% female) were recruited via HRoot (Bock et al., 2014) who mainly were students from local engineering, business and medical schools.<sup>14</sup> The experiment was programmed using z-Tree (Fischbacher, 2007).

Upon arrival, participants randomly drew a ticket from an opaque bag which

---

<sup>11</sup>This was done to avoid placing the investors in a situation in which they would wonder why other participants were typing while they were not.

<sup>12</sup>We needed to run only 5 sessions for Oath-NoCK-Async treatment because there was no pair requirement, given there was only the role of investors.

<sup>13</sup>The number of observations is based on the following power calculation: assuming Type-I error rate of 0.05 and a medium effect size (Cohen's  $d = 0.60$ ), targeting 77 observations per role and per treatment will allow us to uncover the hypothesized effect of oath using a two-tailed Mann-Whitney ranksum test and achieve a power level of 0.95.

<sup>14</sup>See Table 1.B.1 in the Appendix for the summary statistics of the socio-demographic characteristics. The statistics are also reported separately for the role of advisors and investors in Table 1.B.2 and 1.B.3, respectively.

assigned them to computer terminals. Instructions were distributed at the beginning of each part and were read aloud. In part 1, participants played the Truthful Announcement Game. In part 2, participants played the 18 periods of the Announcement Game (or 27 periods in the Oath-NoCK-Async treatment). Finally, they answered the final questionnaire (elicitation of beliefs and socio-demographic questions). The average time of a session was 1 hour and 20 minutes.

Participants' earnings consisted of their payoffs from both parts of the experiment: the payoff in the relevant scenario in the randomly chosen period (9th or 10th) in the Truthful Announcement Game, the payoff from the randomly chosen round in the Announcement Game, and the payment associated with the belief elicitation. On average, participants earned 20 Euros (SD = 6.2), including a 7-Euro fixed show-up fee. Earnings were paid by bank transfer.

## 1.4 Conjectures

In this section, we formulate behavioral conjectures regarding the impact of oath on the announcements made by the advisors and how they can be sensitive to the awareness of the investors. Then, we formulate the behavioral conjectures regarding investors' behavior.

Previous studies have shown that an oath which commit individuals to be honest is an effective instrument in reducing lies (Jacquemet et al., 2019) because it increases the moral cost of lying. In other words, because lying under oath involves two moral transgressions (*i.e.*, lying and breaking the oath), lying becomes less attractive. Nonetheless, recent studies have shown a heterogeneous impact of oath with regards to the size of lies. For example, Jacquemet et al. (2021) found that individuals under oath were less likely to report a full lie in a repeated coin-tossing game, while they found no difference in partial lies between individuals under oath and those in the baseline.<sup>15</sup>

In contrast with the literature, we consider the Announcement Game, which is a strategic game. Advisors can misreport the number of stars, choose the size of lie, and the likelihood that the lie can be detected *ex-post* by the investor. Lying should entail a moral cost of lying, for instance, a feeling of guilt towards oneself (*i.e.*, self-image cost) and perception of being honest (*i.e.*, reputation cost) (see Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019). These two components of the moral cost of lying should increase under oath. In addition, in

---

<sup>15</sup>This may also be related to detectability since reporting 10 heads out of 10 coin tosses should lead to a lie detection with almost certainty, while lying just enough maintains deniability.

the context of Announcement Game, detectable lies should entail both components, while deniable lies should not entail reputation cost since they cannot be detected immediately *ex-post* by the investor.<sup>16</sup>

We formulate the first conjecture for advisors as follows:

**Conjecture 1 for Advisors:** (A common-knowledge oath reduces the prevalence of lies) *Compared to the baseline, advisors in the Oath-CK treatment are less likely to lie to the investors. An oath leads to a significant reduction in detectable lies to a larger extent, and deniable lies to a smaller extent. This is based on the reasoning that detectable lies entail a larger moral cost than deniable lies because the latter entail no reputation cost due to plausible deniability.*

Besides the pure moral costs, making oath common knowledge may invoke an additional cost of lying. Lying under oath whilst the investors are aware of it may enhance a feeling of betrayal and/or guilt. In other words, when oath is common knowledge, advisors may be discouraged to lie to avoid betraying investors' trust. This component should further boost to the moral cost of lying of both detectable and deniable lies when oath is common knowledge with the investors. Thus, advisors should be more likely to lie when the common knowledge element is removed.

We formulate the second conjecture for advisors as follows:

**Conjecture 2 for Advisors:** (Removing the common-knowledge increases the prevalence of lies) *Compared to when oath is common knowledge, advisors engage more in both detectable and deniable lies when oath is not common knowledge. This is based on the reasoning that the common knowledge enhances the moral cost of lying in terms of betraying investors' trust and/or guilt towards investors.*

On the side of the investors, given that a common-knowledge oath reduces the prevalence of lies, it should allow investors to experience relatively more honest announcements overall. This could consequently lead to an improvement in the credibility of the announcements made. Thus, we can expect the investment rates, particularly for 2 and 3 stars announcements, to consequently increase after a common-knowledge oath has been introduced in the Announcement Game.

---

<sup>16</sup>Given a stranger matching procedure, it is very unlikely that the investors will uncover deniable lies.

We formulate the first conjecture for investors as follows:

**Conjecture 1 for Investors:** (Investment rates increase under a common-knowledge oath) *Compared to the baseline, the investment rates in high-numbered announcements in Oath-CK treatment are higher. This is based on the reasoning that a lower prevalence of lies under oath improves credibility of announcements made by the advisors.*

Nonetheless, the improvement in the investment rates under a common-knowledge oath may come from two effects. On the one hand, investors might anticipate advisors to announce more honestly simply because they are aware of the oath. In other words, oath may have an informational value in terms of increasing investors' trust. On the other hand, it could be possible that investors invest more because they respond to the honesty of advisors. That is, they react to the experience of honesty created by a common-knowledge oath. Since the effects could go in both directions, we remain open and formulate the second conjecture for investors and its alternative as follows:

**Conjecture 2 for Investors:** (Information about the oath matters) *If knowing about the oath facilitates investors' trust in the announcements, it is expected that the investment rates in the Oath-NoCK-Async treatment should be lower than those in the Oath-CK treatment.*

**Conjecture 2A for Investors:** (Experience under oath matters) *If the experience of honesty with the advisors is the key determinant of investors' decisions in the Oath-CK treatment, it is expected that the investment rates in the Oath-NoCK-Async treatment are similar to those in the Oath-CK treatment.*

## 1.5 Results

The results section is structured in the following way. We begin by showing that a common-knowledge oath has a large impact on both advisors' and investors' behavior. We then explore the reasons behind the change in investors' behavior. Next, we explore whether the common-knowledge element is material to advisors' behavior or if an oath taken without the knowledge of investors is sufficient to generate honest behavior. Finally, we examine the durability of the effect of the oath over time.

In order to analyse the data, we use ranksum tests and tests of proportions as our main statistical tool to evaluate differences across treatments, using one observations per subject.<sup>17</sup> We also confirm our results with regression analyses, controlling for any differences in individual characteristics.

### 1.5.1 The Impact of a Common-Knowledge Oath on Advisors and Investors

#### Advisors' Behavior

The introduction of the common-knowledge oath has a clear and dramatic impact on advisors' behavior. The fraction of untrue announcements goes from 52.3% to 10.8% ( $p < 0.001$ ) after the introduction of the oath, and in parallel, we observe a very significant increase in the fraction of advisors who always tell the truth: it goes from 9.2% in the absence of an oath to 74.0% when the oath is present ( $p < 0.001$ ).

Importantly, the oath impacts both detectable and deniable lies. Detectable lies are lies that happen when an advisor announces 3 stars but observes fewer than 3 stars. In these situations, the investor may identify that a lie has occurred.<sup>18</sup> Deniable lies happen when an advisor announces 2 stars but observes fewer than 2 stars, or announces 1 star but observes 0 stars. In such a situation, Nature's draw will be consistent with the announcement regardless of whether the draw shows a blank card or a star.

Table 1.1 shows the frequency of detectable and deniable lies among advisors. Looking at the population of advisors as a whole, the reduction in lies after an oath has been introduced impacts both detectable and deniable lies: the relative frequency of the former go from 25.2% to 3.9% ( $p < 0.001$ ) and for the latter from 57.4% to 13.8% ( $p < 0.001$ ).<sup>19</sup> The result regarding the reduction of deniable lies is particularly striking as those lies are precisely the ones that are impossible to detect and that are unimpeded by traditional market mechanisms such as reputation (see [Tergiman and Villeval, 2022](#)).

Interestingly, we note that the oath still has a partial effect among those advisors who lie at least once, which account for about 91% and 26% of advisors in the No

---

<sup>17</sup>Unless otherwise specified, we average choices within a participant and use this average choice as our unit of measure.

<sup>18</sup>When fraudulently announcing 3 stars, the chances that the advisor's lie is detected ranges from 100% if the advisor actually observed 0 stars, to 67% if the advisor observed 1 star among the three cards, to 33% if the advisor observed 2 stars among the three cards.

<sup>19</sup>As the "absolute frequency" of the table shows, the same patterns arise if we look at the fraction of advisors who engage in each type of lie at least once.

Table 1.1: Detectable and Deniable Lies in the No Oath and Oath-CK Treatments

	No Oath	Oath-CK	<i>p</i> -value
<b>All Advisors</b>			
% Untrue announcements	52.3%	10.8%	$p < 0.001$
% Advisors who always tell the truth	9.2%	74.0%	$p < 0.001$
<i>Relative frequency</i>			
Detectable lies	25.2%	3.9%	$p < 0.001$
Deniable lies	57.4%	13.8%	$p < 0.001$
<i>Absolute frequency</i>			
Detectable lies	71.1%	11.7%	$p < 0.001$
Deniable lies	86.8%	23.4%	$p < 0.001$
<b>Advisors who lie at least once</b>			
% Untrue announcements	57.5%	41.3%	$p = 0.010$
<i>Relative frequency</i>			
Detectable lies	27.8%	15.0%	$p = 0.006$
Deniable lies	63.2%	53.2%	$p = 0.301$
<i>Absolute frequency</i>			
Detectable lies	78.2%	45.0%	$p = 0.004$
Deniable lies	95.6%	90.0%	$p = 0.333$

*Notes:* Relative frequency statistics show the average frequency with which participants engaged in any particular type of lie. For example, in the No Oath treatment, advisors make detectable lies 25.2% of the time when they observed 0, 1 or 2 stars (*i.e.*, when they actually could). Absolute frequency statistics show the proportion of advisors who make any particular type of lie at least once. For example, in the No Oath treatment, 71.1% of advisors make at least one detectable lie over the course of the game. This table presents data both for all advisors (Upper panel), and for the subset who make at least one lie (Lower panel).



Oath and Oath-CK treatments, respectively. Among this subgroup of advisors, the frequency of lies reduces from 57.5% without the oath to 41.3% with an oath ( $p = 0.010$ ). The difference in frequency is not constant across detectable and deniable lies: we note a sharp reduction in the relative frequency of detectable lies (27.8% versus 15%,  $p = 0.006$ ), while deniable ones remain high and are not different across treatments for this subset of advisors (63.2% versus 53.2%,  $p = 0.301$ ).<sup>20</sup> This suggests that for the 26% of advisors who lie despite taking the oath, the impact of the oath is only partial and targeted those lies that would obviously make plain that not only a lie has occurred, but also, as a consequence, that the oath has been broken. On the other hand, their engagement in deniable lies is no different to the 91% of advisors who lie at least once without any oath.

These analyses support the following result:

**Result 1.** A common-knowledge oath causes a drastic reduction in the prevalence of lies. Compared to those in the baseline, advisors in Oath-CK treatment engage less frequently in both detectable and deniable lies. However, among the advisors who lie at least once, oath only reduces the frequency of detectable lies, while deniable ones remain unaffected. These findings support Conjecture 1 for advisors.

### Investors' Behavior

We now turn to investors' behavior and look at the impact of the oath, both on overall investment, and after receiving 2 or 3 stars announcements. Table 1.2 shows the average investment rate as well as how investors reacted to announcements for various levels of announcements in both the Announcement Game and in the Truthful Announcement Game. It also displays the fractions of good investments (*i.e.*, the proportion of investment made when the announcement was truthful). Comparing behavior across these two games and across treatments allows us to understand how credible various announcements are and how the common-knowledge oath impacts credibility and investment behavior.

There are significantly more investments after 2 and 3 stars announcements under oath than without the oath. Investment rates after 3 stars announcements increase by almost 20 percentage points after the introduction of a common-knowledge oath, going from 77.8% to 97.7% ( $p < 0.001$ ). The corresponding statistics for behavior after 2 stars announcements are 70.5% and 85.8% ( $p = 0.001$ ). However, because announcements are generally honest, we observe fewer such announcements under

---

<sup>20</sup>The same conclusion holds if we turn our attention to absolute frequencies – see bottom panel of Table 1.1.

Table 1.2: Investments in the Announcement and Truthful Announcement Games, by Treatment

Treatments		Investment Rates		Good Investments	
		No Oath	Oath-CK	No Oath	Oath-CK
0 Stars Announced	% Announcement Game	9.5% (21)	1.7% (59)	—	—
	% Truthful Ann. Game	0.0%	2.6%	—	—
1 Star Announced	% Announcement Game	14.5% (54)	16.8% (77)	87.5% (10)	99.2% (30)
	% Truthful Ann. Game	19.7%	16.9%	—	—
2 Stars Announced	% Announcement Game	70.5% (76)	85.8% (77)	50.2% (74)	83.3% (75)
	% Truthful Ann. Game	98.7%	98.7%	—	—
3 Stars Announced	% Announcement Game	77.8% (76)	97.7% (72)	33.2% (73)	83.8% (72)
	% Truthful Ann. Game	100.0%	98.7%	—	—
Average Inv. Rate/Good Inv.		65.8%	56.9%	42.8%	83.5%

*Notes:* The table shows the fraction of times participants in the role of investors invested in the Announcement and the Truthful Announcement Games, and the proportion of good investments (*i.e.*, proportion of investments made when the announcement was truthful for each case above 0 stars). Number of participants faced with a particular announcement in parentheses.

oath than in the baseline: the fraction of 2 or 3 stars announcements is 90.1% without the oath, and only 58.2% under oath. As the last row of Table 1.2 shows, overall the average investment rates are lower under oath (65.8% versus 56.9%,  $p = 0.004$ ).<sup>21</sup> This has a direct impact on earnings, both for the advisors and the investors.

Due to the impact of a common-knowledge oath, announcements are more honest and credible, investors avoid investing at 0 stars<sup>22</sup> and, at the same time, are more likely to invest when the number of stars is actually 2 or 3. In other words, they invest “better” overall, which is reflected in a higher average proportion of good investments under a common-knowledge oath.<sup>23</sup> This leads to an increase in the average earnings of investors from 149.7 tokens to 160.0 tokens ( $p = 0.015$ , ranksum test). Note that a common-knowledge oath has an asymmetric impact on

<sup>21</sup>Investment in zero stars announcements are noises possibly due to inattention. This fraction seems high in No Oath treatment because of a high prevalence of lies (fewer instances where investors received zero stars announcements). For two investors, they received one zero stars announcement in which they invested. Removing these noises do not change the results reported in the main text.

<sup>22</sup>Note that investing is profitable, starting from 1 star announcement.

<sup>23</sup>The proportions of good investments (average and conditioned on 2 and 3 stars announcements) in the Oath-CK treatment are higher than those in No Oath treatment ( $p < 0.0001$ , ranksum tests). The difference is not significant for the case of 1 star ( $p = 0.206$ ).

the earnings of advisors and investors: investors invest less on average after the oath is introduced, which leads to a significant drop in the average earnings of advisors – from 161.6 to 143.9 tokens ( $p < 0.001$ , ranksum test).

In summary, the investment levels are lower under oath, but the proportion of “good investments” is higher under oath. This has a direct impact of the earnings of both investors and advisors.

These analyses support the following result:

**Result 2.** The impact of a common-knowledge oath consequently promotes the investment rates, particularly the 2 and 3 stars announcements. Overall, the investment level under oath is lower, primarily because the advisors become relatively more honest, which increases the proportion of good investments. This consequently leads to an increase in the earnings for investors. These findings support Conjecture 1 for investors.

### 1.5.2 Understanding Investors’ Behavior in the Oath-CK Treatment

The previous section has shown that the common-knowledge oath reduces the prevalence of lies and has a strong impact on investors’ behavior. We now ask the following question: is this change in behavior due to the investors anticipating that the advisors will be honest, or is it coming from the fact that the investors experience honest announcements? In order to disentangle these two possibilities, we turn to the Oath-NoCK-Async treatment, where all participants took the role of the investors and they were faced with the announcements from the advisors in the Oath-CK treatment (which were predominantly honest), without being aware that those advisors had sworn a common-knowledge oath. If investment rates are similar across the two treatments, it means that it may not be a problem if the oath is not common knowledge *provided that* the advisors are sufficiently honest (*i.e.*, they feel constrained by the oath). In other words, the experience of honesty is a key factor in determining investors’ decision. However, if we find that investment rates are lower in the Oath-NoCK-Async treatment compared to those in the Oath-CK treatment, it means that being aware of the oath facilitates trust, in addition to creating an experience of honesty for the investors.

Table 1.3 places side-by-side the behavior of investors in the Oath-CK and Oath-NoCK-Async treatments. The average investment rates are not different across these two treatments (56.9% versus 55.3%,  $p = 0.890$ ). The gap in investment rates after 0,

Table 1.3: Investment Rates in the Oath-CK and Oath-NoCK-Async treatments

	Oath-CK	Oath-NoCK-Async	<i>p</i> -value
Average Investment Rate	56.9%	55.3%	$p = 0.890$
0 Stars Announced	1.7% (59)	1.1% (76)	$p = 1.000$
1 Star Announced	16.8% (77)	13.0% (77)	$p = 0.747$
2 Stars Announced	85.8% (77)	82.6% (77)	$p = 0.505$
3 Stars Announced	97.7% (72)	93.2% (76)	$p = 0.074$

*Notes:* The table shows the fraction of times participants in the role of investors invested in the Announcement Games. Number of participants faced with a particular announcement in parentheses. For Oath-NoCK-Async treatment, if we restrict the analysis to the first 18 periods (in order to match the number of periods from the Oath-CK treatment), the results correspond to the following figures: Average Investment Rate becomes 54.8%; investment conditioned on 0, 1, 2 and 3 stars become 0.5%, 14.2%, 82.2% and 92.1%, respectively. Results with a restriction are statistically no different than those without restriction.

1, 2 or 3 stars announcements is always small in magnitude (for example, it is only 3.2 percentage points when comparing investment rates after 2 stars announcements), and not statistically different across treatments. We note that the exception to this is the average investment after 3 stars announcements: it is 97.7% under a common-knowledge oath and 93.2% when the investors do not know about the oath. While they are very close to each other, they are nonetheless statistically different, although only marginally so ( $p = 0.074$ ).<sup>24</sup>

From these data, we conclude that making the oath common knowledge is not a *necessary* condition for its impact to be effective on investors: they can indeed learn from experience whether or not the announcements can be trusted. This is consistent with the experience being the main driver of investors' behavior in the Oath-CK treatment, while the awareness of an oath may play a small role in facilitating investors' decision.

These analyses support the following result:

<sup>24</sup>See Table 1.B.7 in the Appendix, which reports the coefficients from the panel Probit regressions (random-effects) on the probability to invest. Overall, the results confirm that investors in the Oath-CK and Oath-NoCK-Async are more likely to invest in 2 and 3 stars announcements. The coefficients of the treatment dummies are not significantly different across all model specifications, which is consistent with the experience explanation.

**Result 3.** Controlling for the empirical distribution of the announcements, investment rates are similar across the Oath-CK and Oath-NoCK-Async treatments. This means that it is not necessary for investors to know about the oath – what matters for them is the experience of honesty, brought about by the introduction of the common-knowledge oath. This finding supports Conjecture 2A against Conjecture 2 for investors.

### 1.5.3 Lying under a No-Common-Knowledge Oath

We established in Section 1.5.1 that a common-knowledge oath had a strong impact on advisors and led to a large reduction in both detectable and deniable lies, and, as a consequence, on lying overall. This improved the investment rates under oath and eventually the earnings of investors. In Section 1.5.2, we showed that for investors, it is not the oath *per se* that mattered (via an anticipation of more honest behavior on part of the advisors), but rather the fact that the investors experienced more honest announcements. In short, the informational value of the oath is small and it is not necessary for investors to be aware of the oath to capture its benefits. Nonetheless, it is the experience of honesty induced by a common-knowledge oath that drives investors' decision.

In this section, we turn to a different question: is the oath *per se* that leads to more honest behavior on part of the advisors, or is the common-knowledge element a necessary condition for their honesty? In order to answer this, we compare advisors' behavior in the Oath-CK and Oath-NoCK treatments, which are summarized in Table 1.4.

Recall that the only difference between the Oath-CK and the Oath-NoCK treatments is that in the latter advisors swore the honesty oath without it being common knowledge with the investors. The difference in advisors' behavior across the two treatments is striking. We find that when the oath is not common knowledge, the fraction of truth-tellers drops by almost half from 74.0% to 39.7% ( $p < 0.001$ ). The top panel shows that when the oath is not common knowledge, the increase in lying impacts both detectable and deniable lies. In fact, the prevalence of lies close to triple across all categories. The relative frequencies of detectable lies go up from 3.9% to 9.6% ( $p = 0.002$ ), while those of deniable lies go up from 13.8% to 37.9% ( $p < 0.001$ ). The pattern is similar if one looks at the absolute frequencies. When looking at those advisors who lied at least once, whether or not the oath was common-knowledge has no impact: behavior is indistinguishable across the two treatments.

Table 1.4: Detectable and Deniable Lies in the Oath-CK and Oath-NoCK Treatments

	Oath-CK	Oath-NoCK	p-value
<b>All Advisors</b>			
% Untrue announcements	10.8%	27.7%	$p < 0.001$
% Advisors who always tell the truth	74.0%	39.7%	$p < 0.001$
<i>Relative frequency</i>			
Detectable lies	3.9%	9.6%	$p = 0.002$
Deniable lies	13.8%	37.9%	$p < 0.001$
<i>Absolute frequency</i>			
Detectable lies	11.7%	32.9%	$p = 0.001$
Deniable lies	23.4%	56.2%	$p < 0.001$
<b>Advisors who lie at least once</b>			
% Untrue announcements	41.3%	45.9%	$p = 0.396$
<i>Relative frequency</i>			
Detectable lies	15.0%	15.8%	$p = 0.593$
Deniable lies	53.2%	62.8%	$p = 0.315$
<i>Absolute frequency</i>			
Detectable lies	45.0%	54.9%	$p = 0.478$
Deniable lies	90.0%	93.1%	$p = 0.660$

*Notes:* Relative frequency statistics show the average frequency with which participants engaged in any particular type of lie. For example, in the Oath-CK treatment, advisors make detectable lies 3.9% of the time when they observed 0, 1 or 2 stars (*i.e.*, when they actually could). Absolute frequency statistics show the proportion of advisors who make any particular type of lie at least once. For example, in the Oath-CK treatment, 11.7% of advisors make at least one detectable lie over the course of the game. This table presents data both for all advisors (Upper panel), and for the subset who make at least one lie (Lower panel).

So, what is an oath worth? Comparing behavior with the No Oath treatment, we can establish that the impact of a no-common-knowledge oath is not insignificant. Given that without an oath, the fraction of truth-tellers was 9.2%, we see that the no-common-knowledge oath is fully impactful for 30.5% of the population (that is, the fraction of truth-tellers in Oath-NoCK minus 9.2% who are truthful anyway without an oath). By comparing further the behavior under a common-knowledge oath, another 34.3% of the population (that is, the fraction of truth-tellers in Oath-CK minus the 30.5% and 9.2%) needs the *common-knowledge element* to be pushed to tell the truth.

Because our design uses random re-matching after every period, direct punishment by the investor onto the advisor is unlikely. Thus, the increase in honest behavior in the Oath-CK treatment relative to the Oath-NoCK treatment is likely due to an increase in intrinsic costs of lying when information about the oath is symmetric. That is, when investors are aware about the oath, advisors are discouraged to tell a lie to avoid feeling guilty towards the investors and/or betraying investors' expectations. The common knowledge about the oath thus enhances the reputational dimension of the moral cost of lying under oath.

This indicates that while the common-knowledge element *per se* is not crucial for investors (as seen in the Section 1.5.2), it is material to the decisions that advisors make. Thus, an oath without common knowledge to the investors would lead to lower levels of honesty and worse outcomes for investors, compared to one with common knowledge. This is reflected in a reduction in investors' earnings when the common knowledge is removed (159.89 tokens in Oath-CK vs. 150.13 tokens in Oath-NoCK,  $p = 0.054$ ).<sup>25</sup>

These analyses support the following result:

**Result 4.** A lack of common knowledge about the oath between advisors and investors leads to an increase in the prevalence of both detectable and deniable lies. This suggests that the common-knowledge element is crucial for advisors' lying strategies. This is consistent with Conjecture 2 for advisors.

---

<sup>25</sup>Considering the joint earnings of advisors and investors, an oath without common knowledge leads to worse outcomes for both parties. Compared to the baseline, introducing oath with common knowledge does not affect the joint earnings (311.25 tokens in No Oath vs. 303.74 tokens in Oath-CK,  $p = 0.632$ ). This is because the investors benefit from the common-knowledge oath while advisors are mostly honest, resulting in a shift in resources. Pooling the data of No Oath and Oath-CK treatments together, oath without common knowledge caused a significant drop in the joint earnings from 307.47 tokens in pooled data to 295.20 tokens in Oath-NoCK ( $p = 0.044$ ). See Table 1.B.5 in the Appendix for earnings of advisors and investors by treatment.

### 1.5.4 The Effect of Oath Over Time

This section investigates whether the effect of oaths on advisors' announcement strategies is durable over the course of the Announcement Game.

Overall, there are no strong time trends in detectable and deniable lies. Figure 1.C.1 and 1.C.2 in the Appendix plot respectively the frequencies of detectable and deniable lies period-by-period across treatments. The fractions of detectable lies in No Oath treatment are high, while those in Oath-CK and Oath-NoCK treatments remain low throughout the Announcement Game. For deniable lies, however, there seems to be a weak increasing trend in Oath-CK (from about 5% in the first period to 20.5% in the last period) and Oath-NoCK treatments (from 20% to 46%). The fraction of deniable lies in No Oath treatment always remains high. We thus focus in more detail the trends of deniable lies over time.

To examine the effect of oath on deniable lies over time, Table 1.5 shows the coefficients from the random-effects Probit regressions where the dependent variable is a dummy indicating that the participant made a deniable lie (coded one if the lie was deniable, zero otherwise), conditioned on observing 0 and 1 star. The independent variables of model 1 include treatment dummies (No Oath treatment as reference category), period, gender (coded one for male, zero otherwise), age (in years) and school (coded one for business school, zero otherwise). Model 2 further controls for the share of non-investments up to that period. Finally, model 3 included the interaction terms of treatment dummies and period. The same set of regressions are done for the subset of advisors who lied at least once (see model 4-6).

We first looking at the results for all advisors (model 1 to 3), all of which indicate that the coefficients of Oath-CK and Oath-NoCK treatments are negative and significant.<sup>26</sup> This confirms our findings on the effect of oath on the prevalence of deniable lies, and that the common-knowledge element further reduces it. Second, in model 1, the coefficient of Period is positive and significant, suggesting that participants were more likely to engage in deniable lies over time. Third, model 2 suggests that it does not seem to matter if investment was received in previous periods or not. This means that a rising engagement in deniable lies is driven simply by the decay of the effect of oath over time. Finally, when controlled for the heterogeneous time trend for each treatment, the variable Period is no longer significant. This is consistent with deniable lies being stable in the No Oath treatment (see Figure 1.C.2 in Appendix). The coefficients of the interaction term between the treatment dummy and Period are positive for both Oath-CK and Oath-NoCK treatments but

---

<sup>26</sup>The difference between the coefficients of Oath-CK and Oath-NoCK treatments is significant for model 1 to 3; all  $p < 0.001$ , Wald tests.



Table 1.5: Probability of Making Deniable Lies

Dep. Var.	(1)	(2)	(3)	(4)	(5)	(6)
Deniable Lie	All Advisors			Lied Once		
No Oath	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Oath-CK	-2.822*** (0.362)	-2.881*** (0.400)	-3.587*** (0.580)	-0.314 (0.335)	-0.325 (0.339)	-1.054** (0.537)
Oath-NoCK	-0.978*** (0.360)	-0.990** (0.392)	-1.531*** (0.493)	0.076 (0.238)	0.096 (0.253)	-0.487 (0.375)
Period	0.043*** (0.012)	0.027** (0.014)	-0.001 (0.022)	0.044*** (0.012)	0.027* (0.014)	-0.002 (0.020)
Male	0.788** (0.314)	0.749** (0.322)	0.780** (0.331)	0.169 (0.223)	0.105 (0.235)	0.124 (0.240)
Age	0.068 (0.042)	0.073* (0.041)	0.072* (0.042)	0.038 (0.034)	0.038 (0.034)	0.036 (0.035)
Business School	0.008 (0.288)	-0.043 (0.295)	-0.040 (0.302)	0.160 (0.220)	0.128 (0.229)	0.143 (0.233)
Share of non-investments	- -	-0.558 (0.399)	-0.424 (0.409)	- -	-0.359 (0.386)	-0.187 (0.389)
Oath-CK X Period	- -	- -	0.062* (0.036)	- -	- -	0.073* (0.039)
Oath-NCK X Period	- -	- -	0.050* (0.030)	- -	- -	0.058* (0.030)
Constant	-2.113** (1.010)	-1.718* (1.038)	-1.492 (1.063)	-0.954 (0.788)	-0.566 (0.819)	-0.297 (0.853)
Log pseudolikelihood	-749.13	-692.68	-688.46	-644.96	-592.08	-586.60
Number of observations	2039	1922	1922	1179	1104	1104
<i>N</i>	226	226	226	133	133	133

*Notes:* This table presents the coefficients from the panel Probit regressions (random-effects) of making deniable lies, conditioned on observing 0 and 1 star for all advisors (model 1 to 3) and those who lied at least once (model 4 to 6). Independent variables in model 1 include treatment dummies (No Oath treatment as reference category), period, gender (coded one for male, zero otherwise), age (in years) and school (coded one for business school, zero otherwise). Model 2 controls for the share of non-investments up to that period ('Share of non-investments'). Model 3 further includes interaction terms between the treatment dummies and period. Model 4-6 correspond to 1-3 respectively for the subset of advisors who lied at least once. Standard errors are clustered at the individual level.

their significance is only marginal. This is consistent with the weak increasing trends of deniable lies under oath over time (see Figure 1.C.2 in the Appendix).

The results for the subset of advisors who lied at least once (model 4-6) largely confirm the results from the non-parametric tests reported in Table 1.1 and 1.4 (that is, the effect of oath on deniable lies is limited when looking at people who lie despite the oath). Model 6, however, shows an interesting finding: after controlling for the heterogeneous time trends for each treatment, it seems that there was an effect of a common-knowledge oath on deniable lies even for this subset of advisors, but as the effect of oath likely decayed over time, it might have offset the initial impact of the oath for this subgroup.<sup>27</sup>

These analyses support the following result:

**Result 5.** While the effect of oath against detectable lies is stable over time, there is an indication that its impact against deniable lies decays over time, although marginally so.

## 1.6 Discussion and Conclusion

Inspired by bankers' oath, which has been implemented in the Netherlands since 2015 and Belgium since 2019, this chapter investigated how such an intervention affects participants in the financial market using a laboratory experiment. Notably, how does an oath affect financial advisors' behavior? Does the oath induce truth-telling? Or does it simply change the way they lie to the investors? In addition, what are the consequences of an oath on the investors? Does an awareness about the oath facilitate trust in advisors? Or is it the investors' past experience that dictate their decision-makings? Our novel design varied the common knowledge element to disentangle between the two mechanisms behind investors' behavior.

Our findings reveal a large impact of a common-knowledge oath on all types of

---

<sup>27</sup>See Table 1.B.6 in the Appendix for the same analysis for detectable lies (for all advisors and the subset who lied at least once). The results confirm the stability of detectable lies across time in all treatments. In addition, they confirm the results of the non-parametric tests reported in Table 1.1 and 1.4: model 1 - 3 show that oath reduces detectable lies, and a common-knowledge element further enhances its impact (the coefficients of Oath-CK and Oath-NoCK treatments are significantly different for all model 1 to 3,  $p < 0.001$ , Wald tests.). For the subset of advisors who lied at least once (model 4 - 6), oaths reduce the prevalence of detectable lies compared to the baseline without oath. The coefficients of Oath-CK and Oath-NoCK treatments are not significantly different for all model 4 to 6;  $p = 0.729$  for model 4,  $p = 0.768$  for model 5, and  $p = 0.370$  for model 6, Wald tests.

lies. Compared to the No Oath treatment, participants in the role of advisors in the Oath-CK engage far less in both detectable and deniable lies. After the introduction of a common-knowledge oath, the fraction of advisors who make detectable lies drop from 71% to 12%, while those that make deniable lies go from 87% to 23%. The dramatic reduction in the prevalence of lies consequently help with the credibility in the announcements, particularly those with 2 and 3 stars, leading to higher level of investments.

Nonetheless, an increase in the level of investment under a common-knowledge oath may be the results of two effects. First, investors may anticipate more honest behavior from advisors who have taken an oath. In other words, oath may facilitate trust in the announcements of the advisors. Second, investors may simply react to their experience since most of the advisors under a common-knowledge oath are predominantly honest. This would imply that experience dictates investment decisions, while information about the oath matter to a smaller extent, at least, for the investors.

Our data are consistent with the experience channel. Largely speaking, it is not the awareness of the oath that drives investment decisions. Rather, it is the honest announcements made under the introduction of a common-knowledge oath that the investors react to by investing more. Simply put, investors have the capacity to learn and deduce from their past experience of honest announcements, *without* having to necessarily know that those announcements had been made under oath. Nonetheless, while an awareness via common knowledge is not a necessary condition for investors to invest, we documented that it is a *necessary* one for advisors as they engage in more detectable as well as deniable lies when oath is not common knowledge. This in turn compromises investment rates. In short, the common-knowledge element is necessary because an oath without a common knowledge does not create sufficient level of honesty required to maintain trust in the market.

An interesting point to note relates to the durability of the effect of an oath. Our data suggest that while its durability against detectable lies is stable over time, it seems that there is a *weak* increasing tendency for advisors who have taken an oath to engage in deniable lies over time. This is not driven by the lack of investment in the past periods due to the inability to attract investments under truth-telling. It suggests however that the decay of oath could be quicker when it comes to a lie that is not detectable outright, compared to those that immediately exposes an individual as a liar (and an oath breaker). This can be explored in the future.

While we need to remain cautious about the external validity of these findings in a more complex environment, the findings presented in this chapter provide important implications regarding the implementation of an oath intervention or other

forms of compulsory code of conduct in the financial institutions. Such a policy intervention needs to be coupled with a well-planned public-relations strategy to ensure a wide coverage of public awareness. In addition, it is extremely crucial for financial institutions to remind employees of the oath, and that investors and the public at large are aware of it as a lack of common knowledge can lead to a Pareto-inefficient outcome – one where advisors, despite an oath, are not sufficiently honest and thus both parties cannot benefit from the intervention. Therefore, an oath intervention with common knowledge brings forth honesty and trust, both of which are the essential foundations of well-functioning markets.

## Appendix

### 1.A Instructions (Translated from French)

#### **No Oath Treatment:**

Hello and welcome to an experiment on decision-making. Please turn off your phone and put it away. You are not allowed to communicate with other participants during the whole experiment, otherwise you will be excluded from the session and any potential earnings.

During this session you can earn money. The amount of money you will earn depends on your decisions and the decisions of other participants in the session. Please read the instructions carefully. All your decisions and responses will be anonymous.

This session consists of two successive parts. The amount you earn at the end of this session is the sum of your earnings in the different parts plus your participation fee of 5 Euros. During the session we will not talk in Euros but in tokens. The conversion rate from tokens to Euros is as follows:

$$100 \text{ tokens} = 4 \text{ Euros}$$

You will be paid via bank transfer by the CNRS. To do this, at the end of the session we will ask for your IBAN number. We will also ask you to send us a bank statement (containing the IBAN) in pdf format with the title “your first name-your last name” to the following e-mail address: gatelab[at]gate.cnrs.fr.

Due to administrative and banking delays, the transfer may take up to approximately two weeks.

We promise that your bank details will only be used to make the transfer. This information will be kept separate from the files containing your decisions and the researchers who process your decisions will not have access to your bank details.

You will now find out the instructions for Part 1. The instructions for Part 2 will be explained at the end of Part 1.

#### **Part 1**

This part consists of 10 periods. The first 8 periods are training rounds and nothing you decide during these 8 periods counts towards determining your actual earnings. For these 8 training periods the earnings stated are therefore hypothetical.

The only periods that can count towards your earnings in this part are the 9th and 10th periods. At the end of the session, the program will draw period 9 or period 10 and the earnings in the drawn period will constitute your earnings for this part. Each of these two periods has 50 chances out of 100 to be drawn.

#### **Description of the task**

In each period, you receive an initial endowment of 100 tokens and you must decide whether to keep these tokens as your earnings for the period or to invest them all in an investment project. If this project is successful, you earn 3 times the number of tokens invested, that is 300 tokens. If the project is not successful, you earn 30 tokens.

**Description of the investment project**

In each period, 3 cards appear on your screen, face down. Each card can have a star (★) or be blank. Each card has 50 chances out of 100 to have a star and 50 chances out of 100 to be blank. These chances are independent for each card.

Thus, your three cards can have a total of zero stars (which happens with 12.5 chances out of 100), a total of 1 star (which happens with 37.5 chances out of 100), a total of 2 stars (which happens with 37.5 chances out of 100), or a total of 3 stars (which happens with 12.5 chances out of 100).

You must press the “Reveal” button to reveal your three cards. In each case, the cards appear in the following format:



After revealing the cards, you must choose whether or not to invest your 100 tokens.

After your choice, the program draws one of your three cards, each card having the same chance of being drawn (so each card has 1 chance out of 3 to be selected).

- If the card drawn has a star and you have invested, the project is a success and you earn 300 tokens (*i.e.*, endowment of 100 – investment of 100 + earnings of 300).
- If the card drawn is blank and you have invested, the project is not a success and you earn 30 tokens (*i.e.*, endowment of 100 – investment of 100 + earnings of 30).
- If you did not invest your tokens, you keep your initial endowment of 100 tokens and therefore earn 100 tokens.

Whatever your choice is, you are informed at the end of the period whether the card drawn by the program from the three cards has a star or not.

Remember that in these 8 training periods, these earnings are hypothetical and nothing you decide in these periods counts towards determining your actual earnings.

To make sure that you have encountered all possible cases, we have in advance chosen cases that correspond to the probabilities announced in the task description.

**Period 9**

If drawn at the end of the session, this period determines your actual earnings for this part. The rules and task are the same as in the previous 8 periods. The only difference is in the way you must make your investment choice.

Your screen will show 3 cards, face down. Each card has 50 chances out of 100 to have a star. In this period, you have to make a decision in each of the following 4 scenarios. Would you invest in the project or not if the program announced that among the three cards there are :

Scenario a) 0 stars?

Scenario b) 1 star?

Scenario c) 2 stars?

Scenario d) 3 stars?

Once you have answered these questions, the program will inform you of the total number of stars among your three cards.

### **Your earnings**

If this period 9 is drawn, your earnings are determined by the answer to the scenario that applies. That is, the one that corresponds to the total number of stars among your three cards. For example, suppose that the three cards hide a total of two stars; in this case, your decision in scenario (c) applies. Another example, suppose the three cards hide a total of three stars; in this case, your decision in scenario (d) applies.

The program then draws one of your three cards at random.

- If you have invested in the project and the card drawn has a star, then you earn 300 tokens (endowment of 100 - investment of 100 + earnings of 300).
- If you have invested in the project and the card drawn is blank, then you win 30 tokens (endowment of 100 - investment of 100 + earnings of 30).
- If you have not invested, you earn the 100 tokens of your initial endowment.

As you can see, the principle is the same as in the 8 training periods, but here you make a decision in each possible scenario.

Since only one of your responses will count towards your earnings if this period is drawn, when you make your decision in each scenario it is in your best interest to treat each scenario as if it were the one that actually counts towards your earnings for this part.

### **Period 10**

If drawn at the end of the session, this period determines your actual earnings for this part. As in the previous periods your screen will show 3 cards, face down. Each card has 50 chances out of 100 to have a star. You must again decide whether or not to invest in the project. However, unlike the previous periods, you only have to make one decision without being informed of the number of cards with a star. Only at the end of the session will you be informed by the programme of the total number of stars among your three cards if this period is selected for payment.

Once you have made your decision, the program will then draw one of the three cards at random.

### **Your earnings**

If period 10 is drawn for payment, your earnings are determined as follows:

- If you have invested and the card drawn has a star, then you earn 300 tokens.
- If you have invested and the card drawn is blank, then you earn 30 tokens.
- If you did not invest, you earn the 100 tokens from your initial endowment.

You will be informed of the period drawn (9 or 10), the card drawn and your earnings in that round at the end of the session.

Please read these instructions again. If you have any questions, please raise your hand or press the red button on the side of your desk. We will come and answer your questions in private immediately.

## Part 2

In this part, each of you will be given a role, either 'A' or 'B'. Half of the participants have a role of A and the other half have a role of B. Your role remains the same for the entirety of part 2: you will never change it.

Part 2 has a minimum of 10 periods and a maximum of 30 periods. The exact number of periods has been decided before the start of the session.

In each period, each of you is paired such that there is one participant A and one participant B in each pair. You will never know the identity of the participant you are paired with. At the beginning of each period, you are randomly re-matched with a new participant. It is unlikely that you will be paired with the same participant two periods in a row.

### Your task in each period

Participant A: Participant A sees three cards on his/her screen, face down. Each card can have a star (★) or be blank. Each card has 50 chances out of 100 to have a star and 50 chances out of 100 to be blank. These chances are independent for each card. Thus, participant A can have a total of 0 stars, 1 star, 2 stars or 3 stars. Participant A can see how many stars he/she has by pressing the "Reveal" button.

Participant A's task is to then announce his/her total number of stars to participant B with whom he/she is matched with. Participant B cannot see Participant A's cards at any time.

Participant B: Participant B receives an initial endowment of 100 tokens and he/she is only informed of the announcement of Participant A with whom he/she is paired about his/her number of stars. Participant B has to decide whether he/she wants to keep his/her tokens or invest them in Participant A's project.

### Determination of earnings

If you are Participant A: You earn a fixed amount of 30 tokens, plus 200 tokens if Participant B has invested in your project.

If you are Participant B: Once you have made your investment decision, the program draws one of Participant A's three cards at random.

- If you have invested in the project and the card drawn has a star, the project is a success and you earn 300 tokens (*i.e.*, endowment of 100 - investment of 100 + earnings of 300).



- If you have invested in the project and the card drawn is blank, the project is not a success and you earn 30 tokens (*i.e.*, endowment of 100 - investment of 100 + earnings of 30).
- If you did not invest, you earn the 100 tokens of your initial endowment.

Whatever your choice is, you are informed at the end of the period whether the card drawn by the program among the three cards has a star or not.

### **Your screen**

At the end of each period, your screen will show a history of what happened in the previous periods. Specifically, you will see four types of information:

1. Your announcements in previous periods, if you are Participant A; or the announcements of the different Participants A you were paired with, if you are Participant B;
2. Whether the card drawn in previous periods had a star or not, regardless of your role and decision;
3. Whether you invested in the previous periods, if you are Participant B; or whether the Participants B you were matched with invested or not, if you are Participant A.
4. Your potential earnings in each previous period.

For each of you, **only one period has already been drawn by the program for payment and it is your decision in that period that will determine your earnings in this part.** Thus, it is in your best interest to make your decision in each period as if it were the period that counts towards your earnings in this part.

Remember that **the pairs are re-matched in each period.**

Please read these instructions again. If you have any questions, please raise your hand or press the red button. We will come and answer your questions in private immediately.

-END OF THE INSTRUCTIONS-

**Oath-CK Treatment:** The printed instructions in Oath-CK are the same as in the No Oath treatment, *except* that there is an additional information about the oath in the instructions for Part 2 as follows.

- - -

### ***Oath***

*At the beginning of part 2, Participants A will have to take an oath about the truthfulness of their announcements in part 2. By this oath, they will swear on their honour that during this experiment they will be honest and always tell the truth.*

- - -

After roles have been assigned, participants in the role of advisors in Oath-CK treatment swore the oath by retyping it via computer. Participants in the role of investors did not type anything during this stage. Instead, they were informed on the screen that those assigned to the role of advisors were taking an oath. After advisors finished taking an oath, all participants were informed that the oath has been taken.

**Oath-No-Common-Knowledge (Oath-NoCK) Treatment:** The printed instructions in Oath-NoCK treatments are the same as in the No Oath treatment. However, participants assigned to the role of advisors in this treatment discovered about the oath via computer screens at the beginning of part 2. During the oath-taking stage, participants in the role of investors typed a neutral sentence: *“The Nobel Prize in Economics for the year 2021 has been awarded to three American economists.”*

**Oath-No-Common-Knowledge-Asynchronous (Oath-NoCK-Async) treatment:** The written instructions for Part 1 of the experiment are the same as in the other treatments. The instructions for part 2 are reported below.

## Part 2

A few weeks ago, participants from the same subjects pool as you took part in an experimental session. These participants performed the same task as you in part 1, under exactly the same conditions as you did. Please read below an extract from the instructions they were given for part 2 (text in italics). The supplementary instructions that only concern you will be given after.

### *Extract from the instructions given to past participants*

*In this part, each of you will be given a role, either ‘A’ or ‘B’. Half of the participants have a role of A and the other half have a role of B. Your role remains the same for the entirety of part 2: you will never change it.*

*Part 2 has a minimum of 10 periods and a maximum of 30 periods. The exact number of periods has been decided before the start of the session.*

*In each period, each of you is paired such that there is one participant A and one participant B in each pair. You will never know the identity of the participant you are paired with. At the beginning of each period, you are randomly re-matched with a new participant. It is unlikely that you will be paired with the same participant two periods in a row.*

### **Your task in each period**

*Participant A: Participant A sees three cards on his/her screen, face down. Each card can have a star (★) or be blank. Each card has 50 chances out of 100 to have a star and 50 chances out of 100 to be blank. These chances are independent for each card. Thus, participant A can have a total of 0 stars, 1 star, 2 stars or 3 stars. Participant A can see how many stars he/she has by pressing the “Reveal” button.*

*Participant A’s task is to then announce his/her total number of stars to participant B with whom he/she is matched with. Participant B cannot see Participant A’s cards at any time.*

*Participant B:* Participant B receives an initial endowment of 100 tokens and he/she is only informed of the announcement of Participant A with whom he/she is paired about his/her number of stars. Participant B has to decide whether he/she wants to keep his/her tokens or invest them in Participant A's project.

**Determination of earnings**

*If you are Participant A:* You earn a fixed amount of 30 tokens, plus 200 tokens if Participant B has invested in your project.

*If you are Participant B:* Once you have made your investment decision, the program draws one of Participant A's three cards at random.

- If you have invested in the project and the card drawn has a star, the project is a success and you earn 300 tokens (i.e., endowment of 100 - investment of 100 + earnings of 300).
- If you have invested in the project and the card drawn is blank, the project is not a success and you earn 30 tokens (i.e., endowment of 100 - investment of 100 + earnings of 30).
- If you did not invest, you earn the 100 tokens of your initial endowment.

Whatever your choice is, you are informed at the end of the period whether the card drawn by the program among the three cards has a star or not.

**Your screen**

At the end of each period, your screen will show a history of what happened in the previous periods. Specifically, you will see four types of information:

1. Your announcements in previous periods, if you are Participant A; or the announcements of the different Participants A you were paired with, if you are Participant B;
2. Whether the card drawn in previous periods had a star or not, regardless of your role and decision;
3. Whether you invested in the previous periods, if you are Participant B; or whether the Participants B you were matched with invested or not, if you are Participant A.
4. Your potential earnings in each previous period.

For each of you, **only one period has already been drawn by the program for payment and it is your decision in that period that will determine your earnings in this part.** Thus, it is in your best interest to make your decision in each period as if it were the period that counts towards your earnings in this part.

Remember that **the pairs are re-matched in each period.**

*End of the extract*

**Supplementary instructions for participants in today's session**

The previous instructions apply to you in this part as well but there are some differences from the previous extract:

1. All participants in today's session are assigned to the role B; you will keep this role throughout the game.
2. Participants in role A are not present in this session: they have already participated in a previous session with other participants in role B.
3. At the beginning of each period, the program will draw a participant A from a previous session with whom you will be matched.
4. Before you decide whether to invest, you will be informed of the announcement made by this participant A from a previous session after the initial draw of his or her three cards.
5. After deciding whether or not to invest, you will be informed of the card that was drawn by the program from the three cards of this participant A from a previous session. This card drawn will determine your earnings for the period if you have decided to invest.
6. Your investment decisions will only affect your earnings and not the earnings of the participants A from previous sessions with whom you are matched.

### Summary

This part has a minimum of 10 periods and a maximum of 30 periods. The exact number of periods has been decided before the start of the session.

In each period you will be informed about the number of stars (0, 1, 2 or 3) announced by a participant A from a previous session in a period drawn at random. You will then have to decide whether or not to invest in the project of participant A from the previous session.

- If you have invested in the project and the card drawn is a star, the project is a success and you earn 300 tokens.
- If you have invested in the project and the card drawn is blank, the project is not a success and you earn 30 tokens.
- If you have not invested, you earn the 100 tokens of your initial endowment.

Whatever your choice is, you are informed at the end of the period whether the card drawn by the program among the three cards has a star or not.

For each of you, **only one period has already been drawn by the program for payment and it is your decision in that period that will determine your earnings in this part.** Thus, it is in your best interest to make your decision in each period as if it were the period that counts towards your earnings in this part.

Remember that **in each period the program randomly selects a decision** (*i.e.*, an announcement made after an initial draw of three cards) **from all the decisions made by participants A in previous sessions. You are re-matched with a new participant A in each period** and it is very unlikely that the program will select the same participant A for you two periods in a row.

Please read these instructions again. If you have any questions, please raise your hand or press the red button. We will come and answer your questions immediately in private.

## 1.B Appendix Tables

Table 1.B.1: Summary Statistics: All Participants

	(1)		(2)		(3)		(4)		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)
	No Oath		Oath-CK		Oath-NoCK		Oath-NoCK-Async							
	Mean	SD	Mean	SD	Mean	SD	Mean	SD						
Age (Years)	21.08	2.94	20.75	1.60	21.17	2.13	21.71	1.85	0.663	0.179	<0.001***	0.054*	<0.001***	0.003***
Male (%)	0.63	0.48	0.41	0.49	0.47	0.50	0.44	0.50	<0.001***	0.004***	0.006***	0.322	0.637	0.730
Business School (%)	0.52	0.50	0.64	0.48	0.44	0.50	0.48	0.50	0.038**	0.159	0.574	0.001***	0.023**	0.547
Observations	152		154		146		77							

*Notes:* This table summarizes the socio-demographic characteristics of participants for each treatment. The *p-value* reported using ranksum test for Age and tests of proportion for Male and Business School. Note that all participants in the Oath-NoCK-Async treatment are assigned the role of investors, while in the other three treatments, participants were randomly assigned either the role of advisors or investors.

Table 1.B.2: Summary Statistics: Advisors

	(1)		(2)		(3)		(1-2)	(1-3)	(2-3)
	No Oath		Oath-CK		Oath-NoCK				
	Mean	SD	Mean	SD	Mean	SD			
Age (Years)	21.34	3.62	20.90	1.84	20.92	2.13	0.625	0.655	0.968
Male (%)	0.66	0.48	0.44	0.50	0.44	0.50	0.007***	0.007***	0.968
Business school (%)	0.50	0.50	0.61	0.49	0.40	0.49	0.169	0.207	0.009***
Observations	76		77		73				

*Notes:* This table summarizes the socio-demographic characteristics of participants in the role of advisors for each treatment (Oath-NoCK-Async treatment omitted as all participants were assigned to the role of investors). The *p-value* reported using ranksum test for Age and tests of proportion for Male and Business School. The proportions of male and students from a business school are unbalanced across treatments. Regression analyzes control for these socio-demographics.

Table 1.B.3: Summary Statistics: Investors

	(1)		(2)		(3)		(4)		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)
	No Oath		Oath-CK		Oath-NoCK		Oath-NoCK-Async							
	Mean	SD	Mean	SD	Mean	SD	Mean	SD						
Age (Years)	20.82	2.03	20.60	1.32	21.42	2.11	21.71	1.85	0.934	0.023**	<0.001***	0.006***	<0.001***	0.1255
Male (%)	0.61	0.49	0.38	0.49	0.49	0.50	0.44	0.50	0.004***	0.169	0.042**	0.150	0.412	0.526
Business School (%)	0.54	0.50	0.66	0.48	0.48	0.50	0.48	0.50	0.120	0.463	0.465	0.023**	0.022**	0.989
Observations	76		77		73		77							

*Notes:* This table summarizes the socio-demographic characteristics of participants in the role of investors for each treatment. The *p-value* reported using ranksum test for Age and tests of proportion for Male and Business School. Regression analyses control for these socio-demographics and they do not affect the decision to invest (See Table 1.B.7).

Table 1.B.4: Investments in the Announcement and Truthful Announcement Games, by Treatment

Treatments		No Oath	Oath-CK	Oath-NoCK	Oath-NoCK-Async
0 Stars Announced	% Announcement Game	9.52% (21)	1.69% (59)	4.0% (50)	1.09% (76)
	% Truthful Ann. Game	0.0%	2.59%	1.36%	0.0%
1 Star Announced	% Announcement Game	14.50% (54)	16.84% (77)	13.02% (70)	12.95% (77)
	% Truthful Ann. Game	19.73%	16.88%	17.80%	11.68%
2 Stars Announced	% Announcement Game	70.51% (76)	85.76% (77)	72.61% (73)	82.57% (77)
	% Truthful Ann. Game	98.68%	98.70%	98.63%	98.70%
3 Stars Announced	% Announcement Game	77.77% (76)	97.70% (72)	89.49% (69)	93.18% (76)
	% Truthful Ann. Game	100.00%	98.70%	98.63%	100.00%

*Notes:* The table shows the fraction of times participants in the role of investors invested in the Announcement and Truthful Announcement Games. Figures are reported at individual level (*i.e.*, of all the times the announcement of ‘x’ star was received, how many times did the investor invest). Number of participants faced with a particular announcement in parentheses. For the Oath-NoCK-Async treatment, restricting behaviors to the first 18 periods yields the following- Announcement Game: 0.48%, 14.23%, 82.16% and 92.12% for 0, 1, 2 and 3 stars, respectively; Truthful Announcement Game: 0%, 11.68%, 98.70% and 100% for 0, 1, 2 and 3 stars, respectively.

Table 1.B.5: Average Earnings of Advisors and Investors, by Treatment

Treatments	No Oath	Oath-CK	Oath-NoCK	Oath-NoCK-Async
Advisors	161.57 (31.06)	143.85 (32.40)	145.06 (37.74)	140.53 (34.98)
Investors	149.67 (31.29)	159.89 (31.16)	150.13 (30.99)	158.46 (25.41)

*Notes:* This table displays the average earnings (in tokens) by advisors and investors by treatment. Standard deviation in parentheses. Figures reported at individual level. Note that the figures of advisors in the Oath-NoCK-Async treatment are simulated based on investors’ decisions. Restricting to the first 18 periods, the figures for Oath-NoCK-Async treatment become 139.66 (38.20) for ‘advisors’ and 157.46 (29.72) for investors.

Table 1.B.6: Probability of Making Detectable Lies

Dep. Var.	(1)	(2)	(3)	(4)	(5)	(6)
Detectable Lie	All Advisors			Lied Once		
No Oath	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Oath-CK	-2.201*** (0.335)	-2.172*** (0.344)	-2.543*** (0.525)	-0.828** (0.409)	-0.805* (0.416)	-1.216** (0.572)
Oath-NoCK	-1.190*** (0.277)	-1.165*** (0.288)	-1.114*** (0.369)	-0.674** (0.272)	-0.669** (0.288)	-0.666* (0.370)
Period	0.008 (0.012)	0.011 (0.013)	0.009 (0.017)	0.009 (0.012)	0.011 (0.013)	0.007 (0.017)
Male	0.696*** (0.252)	0.665*** (0.254)	0.667*** (0.256)	0.361 (0.257)	0.335 (0.265)	0.337 (0.266)
Age	-0.059 (0.038)	-0.066* (0.039)	-0.066* (0.040)	-0.063* (0.035)	-0.072* (0.036)	-0.072* (0.037)
Business School	-0.139 (0.251)	-0.134 (0.255)	-0.140 (0.255)	-0.031 (0.244)	-0.027 (0.253)	-0.032 (0.254)
Share of non-investments	- -	-0.242 (0.374)	-0.266 (0.357)	- -	0.015 (0.382)	0.008 (0.364)
Oath-CK X Period	- -	- -	0.033 (0.036)	- -	- -	0.038 (0.039)
Oath-NCK X Period	- -	- -	-0.005 (0.026)	- -	- -	-0.001 (0.027)
Constant	-0.317 (0.859)	-0.132 (0.873)	-0.110 (0.896)	0.142 (0.828)	0.282 (0.865)	0.323 (0.884)
Log pseudolikelihood	-837.10	-784.09	-783.18	-798.26	-746.71	-745.75
Number of observations	3551	3356	3356	2096	1979	1979
<i>N</i>	226	226	226	133	133	133

*Notes:* This table presents the coefficients from the panel Probit regressions (random-effects) of making detectable lies, conditioned on observing 0 and 1 star for all advisors (model 1 to 3) and those who lied at least once (model 4 to 6). Independent variables in model 1 include treatment dummies (No Oath treatment as reference category), period, gender (coded one for male, zero otherwise), age (in years) and school (coded one for business school, zero otherwise). Model 2 controls for the share of non-investments up to that period ('Share of non-investments'). Model 3 further includes interaction terms between the treatment dummies and period. Model 4-6 correspond to 1-3 respectively for the subset of advisors who lied at least once. Standard errors are clustered at the individual level.

Table 1.B.7: Probability of Investing in 2 and 3 Stars Announcements

Dep. Var. Investment Decision	(1)	(2)	(3)
No Oath	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
Oath-CK	0.972*** (0.187)	0.652*** (0.204)	1.128*** (0.326)
Oath-NoCK	0.231 (0.172)	0.038 (0.181)	0.064 (0.280)
Oath-NoCK-Async	0.894*** (0.200)	0.600*** (0.221)	0.680** (0.329)
Period	-0.017*** (0.005)	<0.001 (0.006)	<0.001 (0.007)
Male	0.196 (0.135)	0.209 (0.137)	0.190 (0.137)
Age	-0.001 (0.035)	-0.004 (0.034)	0.002 (0.035)
Business School	-0.035 (0.135)	-0.071 (0.138)	-0.065 (0.139)
Nb. of lies detected	-	-0.242*** (0.060)	-0.183*** (0.069)
% of blank cards on 2 Stars	-	-0.414** (0.164)	-0.348 (0.295)
Oath-CK X Nb. of lies detected	-	-	-0.453* (0.241)
Oath-NoCK X Nb. of lies detected	-	-	-0.136 (0.117)
Oath-NoCK-Async X Nb. of lies detected	-	-	-0.146 (0.205)
Oath-CK X % of blank cards on 2 Stars	-	-	-0.727 (0.456)
Oath-NoCK X % of blank cards on 2 Stars	-	-	0.174 (0.417)
Oath-NoCK-Async X % of blank cards on 2 Stars	-	-	0.039 (0.461)
Constant	0.866 (0.742)	1.249* (0.747)	1.010 (0.785)
Log pseudolikelihood	-1403.93	-1524.84	-1519.47
Number of observations	4214	3750	3750
<i>N</i>	303	303	303

*Notes:* This table presents the coefficients from the panel Probit regressions (random-effects) of investing, conditioned on receiving 2 and 3 stars announcements (coded one for invest, zero otherwise). Independent variables in model 1 include treatment dummies (No Oath treatment as reference category), period, gender (coded one for male, zero otherwise), age (in years) and school (coded one for business school, zero otherwise). Model 2 controls for the number of lies detected up to the period ('Nb. of lies detected') and the percentage of blank cards on 2 stars announcements up to the period ('% of blank cards on 2 Stars'). Model 3 further includes interaction terms between the treatment dummies and number of lies detected and blank cards on 2 stars. Standard errors are clustered at the individual level. Data in Oath-NoCK-Async treatment are restricted to the first 18 periods to have a balanced panel with the other treatments. Relaxing this restriction does not affect the results.



### 1.C Appendix Figures

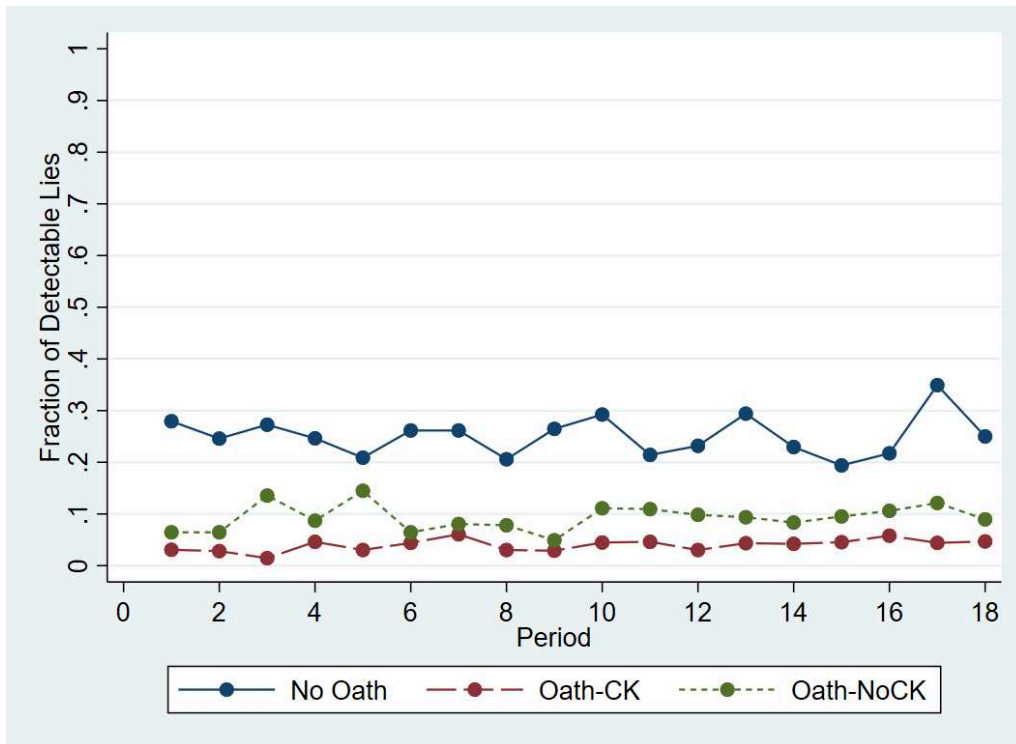


Figure 1.C.1: Fraction of Detectable Lies in Each Period, by Treatment

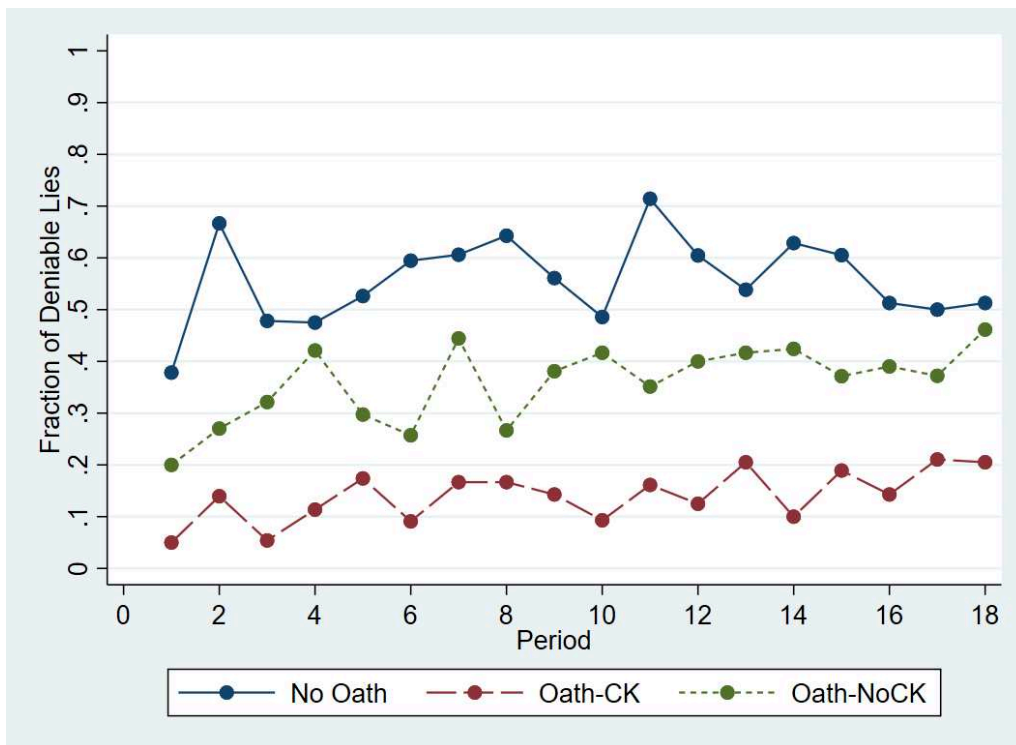


Figure 1.C.2: Fraction of Deniable Lies in Each Period, by Treatment

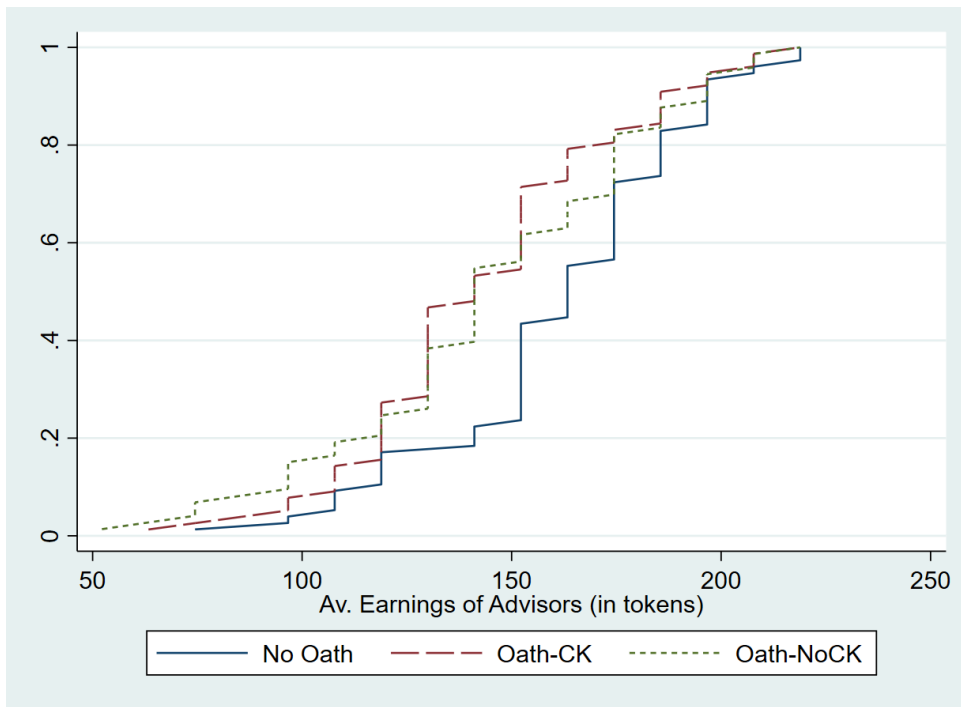


Figure 1.C.3: Cumulative Distribution Function of Advisors' Earnings, by Treatment

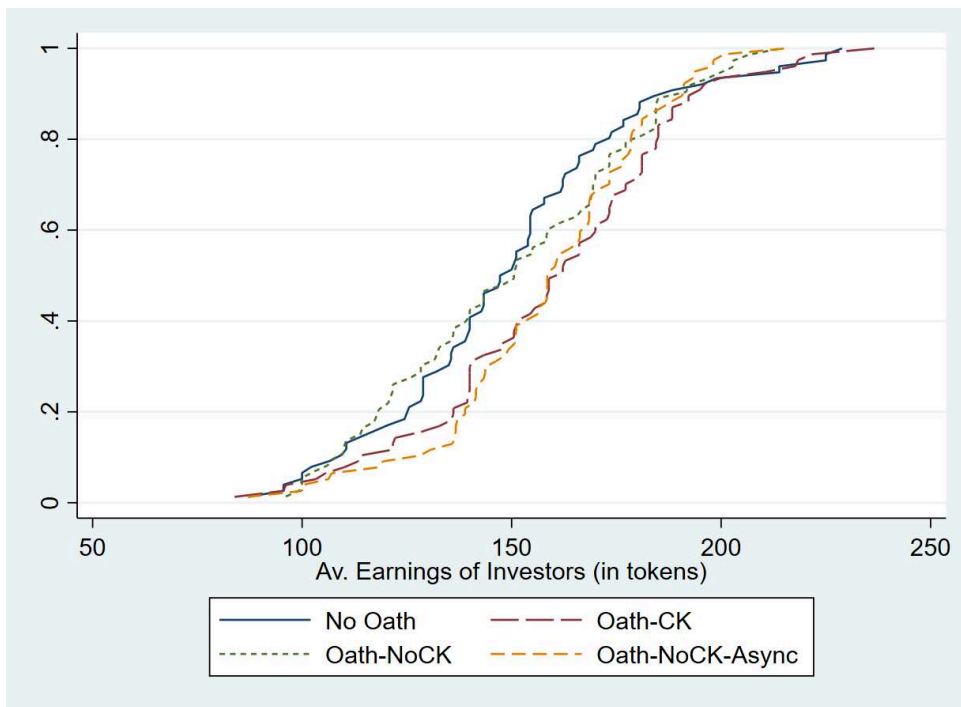


Figure 1.C.4: Cumulative Distribution Function of Investors' Earnings, by Treatment

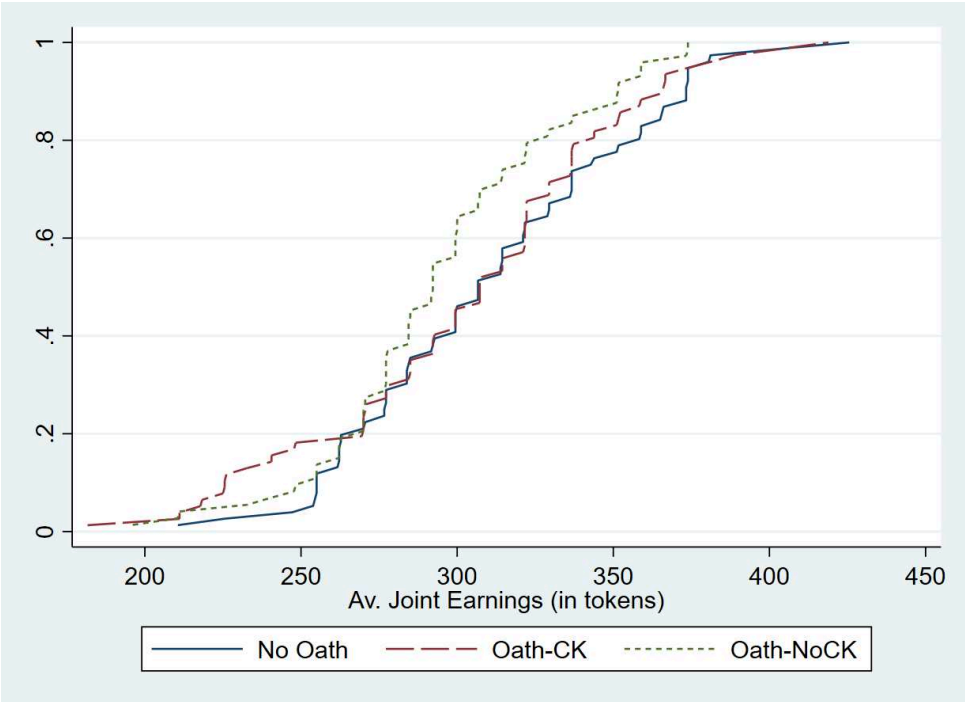


Figure 1.C.5: Cumulative Distribution Function of Joint Earnings of Advisors and Investors, by Treatment

## 1.D Belief Elicitations

This section reports the results of the beliefs elicitation. We elicited beliefs of participants at the end of session. Investors were asked to estimate the proportion of honest announcement for each number of stars made to them during the experiment (*i.e.*, the first order belief about the honesty of advisors). Advisors were asked to estimate the guesses made by the investors about the proportion of honest announcement for each number of stars (*i.e.*, the second order belief of the advisors) and the proportion of honest announcement for each number of stars announced by all other advisors in the session (*i.e.*, the first order belief about advisors). The order of the questions for advisors were randomized. They could earn an additional payoff of 0.50 Euros if their answer in one randomly selected guesses was within an interval of 5% point.

The responses were somewhat noisy. For example, investors should believe that if the advisors had announced 0 stars, it should be true because they have nothing to earn from announcing ‘0’. Thus, ‘100%’ seems to be a rational response for 0 stars announcement, unless the investors believe that advisors can make mistakes. Therefore, we implement the following as an exclusion criterion: if the participant guessed that the proportion of honest 0 stars announcement was below 95%, he/she likely misunderstood the question and thereby remain excluded.

Based on the above criteria, we excluded 56 of 303 investors, 103 of 226 advisors for the first order belief and 117 of 226 advisors for the second order belief. While quite a number of investors pass this criterion, about half of the advisors are excluded. We thus take interpretations of advisors’ beliefs with a grain of salt.

Table 1.D.1: First Order Belief of Investors

	(1)		(2)		(3)		(4)	
	No Oath		Oath-CK		Oath-NoCK		Oath-NoCK-Async	
	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
0 Stars	99.33 (1.75)	15	99.53 (1.40)	43	99.67 (1.17)	37	99.27 (1.77)	55
1 Star	87.20 (26.44)	50	93.85 (10.08)	61	82.93 (23.33)	58	86.40 (17.64)	55
2 Stars	52.92 (21.60)	70	82.57 (17.36)	61	64.51 (22.55)	60	73.74 (21.96)	55
3 Stars	42.94 (30.23)	70	85.57 (24.64)	56	63.26 (31.84)	57	77.87 (23.92)	55

*Notes:* This table displays the first order beliefs of investors (that is, what they think about the honesty of advisors) by treatment. The number of observations varies from case to case depending on the actual received announcements during the Announcement Game.

Table 1.D.1 and 1.D.2 display the first order beliefs of investors and the distance between beliefs and the actual level of honesty of advisors, respectively. It seems that having information about the oath helped investors to learn better about detectable and deniable lies. The distances of beliefs and actual honesty for 2 and 3 stars announcements for each given treatment are similar and not significantly different, except for the Oath-CK treatment (about 6% point,  $p = 0.009$ , signrank test). Despite the same empirical distribution of announcements in the Oath-NoCK-Async treatment, it is plausible that without being aware about the oath, investors suspected the credibility of 3 stars, even though Nature’s draw did not contradict the announcements and/or that an oath may facilitate trust in general. This is likely since investors in Oath-CK treatment got closer to the true level of honesty compared to those in Oath-NoCK-Async treatment by about 5

percentage points, but the significance remains weak ( $p = 0.063$ , ranksum test). However, investors in Oath-CK and Oath-NoCK-Async treatments were no different in their accuracy when it comes to 2 stars announcements ( $p = 0.416$ , ranksum test). This may imply that knowing about the oath does not further help investors to trust 2 stars announcements because of its deniability in nature, but it may have a marginal informational value for 3 stars announcements.

Table 1.D.2: Distance Between Beliefs of Investors and Actual Honesty

	(1)		(2)		(3)		(4)	
	No Oath		Oath-CK		Oath-NoCK		Oath-NoCK-Async	
	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
0 Stars	0.66 (1.75)	15	1.24 (5.20)	43	2.03 (7.95)	37	1.33 (4.74)	55
1 Star	21.40 (35.00)	50	6.14 (10.08)	61	18.13 (25.53)	58	13.99 (17.49)	55
2 Stars	20.96 (13.87)	70	18.45 (15.23)	61	20.32 (13.85)	60	20.04 (13.79)	55
3 Stars	24.54 (20.12)	70	12.69 (21.06)	56	17.30 (17.84)	57	17.18 (19.09)	55

*Notes:* This table displays the distance between the first order beliefs of investors and the actual proportions of honest announcements (that is, how far are their beliefs from the truth) by treatment. The number of observations varies from case to case depending on the actual received announcements during the Announcement Game.

Table 1.D.3 and 1.D.4 display the first and second order beliefs of advisors, respectively. Table 1.D.5 displays the distance between the first order belief of advisors and the actual level of honesty. The distance between belief and actual level of honesty in the Oath-NoCK treatment was higher than that in the Oath-CK and No Oath treatment for 2 stars ( $p = 0.002$  and  $p = 0.081$  respectively, ranksum tests) and 3 stars ( $p = 0.002$  and  $p < 0.001$  respectively, ranksum tests). These high values in Oath-NoCK treatment may reflect uncertainty about what the other advisors might behave under oath without common knowledge. On the other hand, those in No Oath treatment may expect others to lie, while those in Oath-CK treatment expect others to be honest, which might help them get closer to the actual level of honesty.

Table 1.D.3: First Order Belief of Advisors

	(1)		(2)		(3)	
	No Oath		Oath-CK		Oath-NoCK	
	Mean	SD	Mean	SD	Mean	SD
0 Stars	98.40	2.29	99.10	1.94	98.54	2.27
1 Star	90.47	13.64	92.66	14.12	92.81	12.10
2 Stars	42.27	22.55	84.71	21.37	67.11	28.29
3 Stars	33.07	26.55	88.82	17.01	67.43	34.08
N	40		39		44	

*Notes:* This table displays the first order beliefs of advisors (that is, what they think about the honesty of all advisors in that session) by treatment. The number of observations varies from case to case depending on the actual received announcements during the Announcement Game.

Table 1.D.4: Second Order Belief of Advisors

	(1)		(2)		(3)	
	No Oath		Oath-CK		Oath-NoCK	
	Mean	SD	Mean	SD	Mean	SD
0 Stars	98.10	2.4	99.08	1.93	98.55	2.27
1 Star	87.63	16.86	88.37	18.60	86.26	21.37
2 Stars	49.36	22.67	81.64	19.42	60.05	22.80
3 Stars	33.65	25.73	88.18	18.86	58.91	34.17
N	38		37		34	

*Notes:* This table displays the second order beliefs of advisors (that is, what they think investors think about their honesty) by treatment. The number of observations varies from case to case depending on the actual received announcements during the Announcement Game.

Table 1.D.5: Distance Between Beliefs of Advisors and Actual Honesty

	(1)		(2)		(3)	
	No Oath		Oath-CK		Oath-NoCK	
	Mean	SD	Mean	SD	Mean	SD
0 Stars	1.60	2.29	0.97	2.06	3.69	5.00
1 Star	12.50	11.26	7.28	13.99	7.04	11.29
2 Stars	19.46	14.37	15.47	12.27	24.96	14.85
3 Stars	20.47	16.11	12.15	14.37	30.69	17.27
N	40		39		44	

*Notes:* This table displays the distance between the first order beliefs of advisors and the actual proportions of honest announcements (how far are their beliefs from the truth) by treatment. The number of observations varies from case to case depending on the actual received announcements during the Announcement Game.

## 1.E Survey on Prolific

This section reports the survey responses conducted via Prolific with 200 individuals residing in the Netherlands. The survey was collected via Google Forms in Dutch and translated back to English. Responses are reported in the parentheses.

### Consent

This study includes a short survey conducted by Chloe Tergiman (The Pennsylvania State University), Marie Claire Villeval (University of Lyon) and Sorravich Kingsuwankul (University of Lyon). It takes about 3 minutes to complete the survey.

Participation is voluntary. You can withdraw from the survey at any time without giving a reason. This will not affect the relationship between you, the researchers or Prolific in any way. All aspects of the study are confidential and anonymous. We will not ask for your name or any information that may identify you during the study. A report of this survey may be submitted for publication, but all information will be used in an aggregate form.

Upon completion of this survey, you will receive a fixed fee of 1€.

If you have specific questions about the survey, you can contact us by e-mail at [kingsuwankul\[at\]gate.cnrs.fr](mailto:kingsuwankul[at]gate.cnrs.fr)

If you agree with the above terms, please select "I agree" to continue with the survey.

- I agree (*Continue*)
- I do not agree (*Exit survey*)

### Please enter your Prolific ID:

(*Input box here*)

### Do you live in the Netherlands?

- Yes (*Continue*)
- No (*Exit survey*)

Q1. Could you say in general that most people can be trusted or that you have to be very careful when dealing with people?

- Most people can be trusted (*65%*)
- You have to be very careful (*35%*)

Q2. Please indicate on a 5-point scale from 1 "I don't trust at all" to 5 "I trust completely", how much you trust each of these groups (*Likert scales 1 - 7*)

- Doctors (*Mean = 4.05, SD = 0.72*)
- Bankers (*Mean = 2.67, SD = 0.85*)
- Fire department (*Mean = 4.55, SD = 0.58*)
- Banks (*Mean = 2.67, SD = 0.94*)



- Large companies (*Mean = 2.60, SD = 0.81*)

Q3. How often do you seek financial advice from your banker or financial advisor?

- I do not have a financial advisor (*63%*)
- Less than once a year (*25%*)
- Once a year (*9%*)
- A few times a year (*2.5%*)
- Every month or more often (*0.5%*)

Q4. Please select any relevant services that you have recently used/used at your bank. If other, please specify.

- Savings accounts/short-term accounts (*88%*)
- Credit/debit cards (*71%*)
- Insurance (e.g., home, life, health) (*50%*)
- Loans (e.g., personal, car, home) and mortgages (*33.5%*)
- Investment (e.g., stocks, bonds, securities) (*27%*)
- Other (*0.5%*)

Q5. Do you know whether bankers in the Netherlands are required to take an oath of good conduct?

- I'm sure bankers don't have to take an oath (*3%*) (Go to 6B then 7B)
- Bankers unlikely to be required to take an oath (*15%*) (Go to 6B then 7B)
- I don't know (*49.5%*) (Go to 6B then 7B)
- It is likely that bankers will have to swear an oath (*21%*) (Go to 6A then 7A)
- I'm sure bankers have to swear an oath (*11.5%*) (Go to 6A then 7A)

Q6A. Do you think that asking bankers to take an oath of good conduct has a real effect on actual professional behavior and the elimination of malpractice? Answer by choosing a number from 1 to 7, where 1 means "no real impact" and 7 means "of great impact."

(*Likert Scale 1 - 7: Mean = 3.32, SD = 1.62, N=65*)

Q7A. Does it increase your confidence in your bank and/or banker if bank employees are required to take an oath of good conduct?

- Yes, it increases the trust I have in my bank and/or banker (*41.5%*)
- No, it does not increase the trust I have in my bank and/or banker (*52.3%*)
- I don't know (*6.2%*)

Q6B. Do you think that requiring bankers to take an oath of good conduct would have a real effect on actual professional conduct and the elimination of malpractice? Answer by choosing a number from 1 to 7, where 1 means “no real impact” and 7 means “of great impact.”

*(Likert Scale 1 - 7: Mean = 3.22, SD = 1.72, N=135)*

Q7B. Would it increase your trust in your bank and/or banker if bank employees were required to take an oath of good conduct?

- Yes, it would increase the trust I have in my bank and/or banker (31.9%)
- No, it would not increase the trust I have in my bank and/or banker (57.8%)
- I don't know (10.4%)

Q8. What do you think about the effectiveness of this banking oath?

- Not effective at all (22.5%)
- Moderately effective (57.5%)
- Highly effective (6.5%)
- I don't know (13.5%)

Q9. What is your gender?

- Male (50.5%)
- Female (48.5%)
- Other (1%)

Q10. In what year were you born? (*e.g.* 1970)

*(Mean age = 36.34, SD = 9.52)*

Q11. What is the highest level of education you have obtained? If you are currently enrolled, please tick the highest level of education you have attained.

- High school or below (11.5%)
- Bachelor (52.5%)
- Master or above (36%)

Q12. What is your current status?

- Student (10.5%)
- Employee (63%)
- Self-employed (21%)
- Retired (0.5%)
- Unemployed (5%)

Q13. What is your annual personal net income (your salary after income tax, social security contributions and pension contributions have been deducted)?

- €<14999 (23%)
- €15000 – €29999 (26.5%)
- €30000 – €44999 (29.5%)
- €45000 – €59999 (15%)
- €60000 and more (6%)

Q14. The next question is about the following problem. In questionnaires like ours, some participants sometimes do not read the questions carefully and just click quickly on the questionnaire. This means that there are many random answers that jeopardize the results of surveys. To show that you have read our questions carefully, give “bankers” as your answer to the following question. What is the profession indicated?

- Firefighters (0%)
- Bankers (100%)
- Doctors (0%)
- Teachers (0%)
- Police officers (0%)

**End of the survey**

You have completed the survey. We thank you for your time.

Important: Please complete the following 3 steps to register your survey response and receive your payment.

1. Write down this survey code: (*Survey code here*)
2. Click on ‘Submit’ on this page to register your responses. If you do not complete this step, we will not receive your information and cannot reward you.
3. Enter the survey code into your Prolific account to register your submission.

# Chapter 2

## Individual Oath-Swearing and Lying in Group

### 2.1 Introduction

Frauds and financial scandals often involve groups of individuals who are engaged in unethical collaboration. Take for an example the Wells Fargo scandal. The incentive structure and goals set by the top executives pressurized employees over the edge, to the point that they had to engage in illegal practices to meet the ‘unrealistic sale quotas.’ This included creating fake accounts and pushing unnecessary financial products in order to inflate the firm’s performance, and thereby the stock value. Such stories have led to a decline in public trust in the banking industry since the aftermath of the global financial crisis in 2008 ([Guiso et al., 2008](#); [Sapienza and Zingales, 2012](#)). To this end, a bankers’ oath (similar to the Hippocratic Oath in medicine) has been proposed with the aim of combating professional malpractice and restoring public trust. The Netherlands made bankers’ oath legally binding in 2015, followed by Belgium in 2019.

Evidence in psychology and behavioral economics show that oath reduces lying at an individual level (see [Heinicke et al., 2019](#); [Jacquemet et al., 2019, 2020, 2021](#)) because it increases the moral cost of lying. This can be a cost-effective intervention to promote individual honesty. However, an important unanswered question is how would such an oath work against dishonesty in a group setting? On the one hand, an oath increases the intrinsic cost of lying and makes the honesty norm more salient. This can encourage individuals to promote ethical behavior within their group. On the other hand, group communication permits exchanges of justifications, which can compromise its impact. Will the impact of oath against lying extend to a group setting?

It is interesting to examine this phenomenon for two reasons. First, previous studies have found that groups are generally more dishonest than individuals (*e.g.*, [Conrads et al., 2013](#); [Weisel and Shalvi, 2015](#); [Kocher et al., 2018](#)). Thus, it puts the honesty oath to a stress test in an environment where the prevalence of lying is high. Second, individuals can typically communicate with other members of their group, which allows them to learn about others' behaviors. This can in turn influence their own behavior, giving rise to peer conformity. Given that studies on peer effects and lying showed that people tend to lie more after observing the norm violations of others (see [Gino et al., 2009](#); [Fosgaard et al., 2013](#); [Innes and Mitra, 2013](#); [Kroher and Wolbring, 2015](#); [Diekmann et al., 2015](#)), investigating whether an honesty oath can limit lying in a setting where individuals tend to conform to others is an interesting research avenue.

Additionally, this chapter tests whether the impact of an honesty oath can interact with the incentive structure group members face. In particular, it explores the case of the payoff commonality rule, proposed by [Kocher et al. \(2018\)](#), representing different organizational structures. Under this incentive scheme, group members are required to coordinate on an action, and failure to do so results in a loss for the whole group. On the other hand, the payoffs of the group members are independent under no payoff commonality. [Kocher et al. \(2018\)](#) found that individuals in a group setting lied to a similar extent under the two incentive schemes because they could communicate, exchange justifications, and thereby conformed to each other. Nonetheless, these incentive schemes may matter under oath because the payoff commonality rule, in addition to pure conformity effect, exerts a pressure on individuals to lie in a group setting. Simply put, a person in a group that earns a common payoff may lie not only out of pure conformity (*i.e.*, I do what others are doing) but also out of peer pressure from the group members (*i.e.*, I do it because others want me to). As the payoff commonality rule creates peer pressure on individuals to conform to the group, it may crowd out the impact of an honesty oath. Therefore, one may expect a larger impact of oath when group members do not earn a common payoff.

The experiment uses a variant of the observed cheating game developed by [Gneezy et al. \(2018\)](#) in a mixed design (similar to that of [Kocher et al., 2018](#)). There are three parts (within-subjects). In part 1, participants play the game in an individual setting, without any social interactions. A video of a die outcome is shown to the participants via computer. They are asked to memorize and report it later to obtain a payoff. In this game, participants can misreport the outcome of the die and earn additional payoffs. Since the die outcome is observable by the experimenter, lying behavior can be identified *ex-post* at an individual level. The deci-

sion in part 1 is used as a proxy for individual moral type (honest or dishonest). In part 2, participants play the same game but in a group setting. They are randomly matched to form a group of three. All the three group members first view the same video of a die outcome, after which they can anonymously chat via computer for five minutes. The chat is free-form and can be about anything, except revealing personal information. After the chat, group members report the number *individually* as in part 1. The decision in part 2 will be of main interest in this chapter, which captures lying behavior in a group setting. Finally, in part 3, participants play the game in an individual setting as in part 1. Since participants have interacted with two other group members in part 2, their decision in part 3 will reveal how group interactions affect their individual honesty.

There are four (2x2) between-subjects treatments varying in the incentive schemes and the existence of oath. The between-subjects dimension is introduced at the beginning of part 2 of the experiment. This means that part 1 remains the same across treatments. In **BaseNoPC**, payoffs of the group members are independent. Participants are informed that the number they report will not affect the payoff of their group members, and vice versa. In **BasePC**, group members face payoff commonality. They are informed that all the three members must report the same number to earn the payoff associated with the number. Otherwise, they earn nothing. BaseNoPC and BasePC closely follow [Kocher et al. \(2018\)](#) and serve as the baseline conditions where individuals lie in a group setting under two different incentive structures. The two remaining treatments, **OathNoPC** and **OathPC**, follow the payoff commonality rule of their respective baseline, *except* that at the beginning of Part 2 (before receiving the instructions), all participants swear an honesty oath in which they commit to be honest and will tell the truth for *the rest of the experiment*, and it is common knowledge.

Results in the baseline conditions replicate the *dishonesty shift in group* of [Kocher et al. \(2018\)](#). Individuals lied more in a group setting (part 2) than in an individual setting (part 1) – the fractions of liars in BaseNoPC and BasePC drastically increased from 57.3% and 56.9% in part 1 to 91.4% and 94.9% in part 2, respectively. In addition, the lying rates in a group setting are similar across the baselines, replicating the conformity effect reported in [Kocher et al. \(2018\)](#).

Comparing lying behavior in part 2 across the treatments reveals that the impact of an honesty oath extends to a group setting but it depends on the incentive structure – oath reduces lying in group *only* under no payoff commonality. When group members do not earn a common payoff, the fraction of liars reduces from 91.4% to 76.5%. The reduction in the lying rates caused by an oath is marginal under payoff commonality – from 94.9% to 86.2%. Thus, the impact of oath against

lying in group is larger under no payoff commonality. This is in line with the notion that the payoff commonality rule can exert pressure on individuals to conform to the group, crowding out the impact of the oath on honest reporting.

The analysis of lying behaviors in part 2 and 3 shed light on the reasons behind the larger impact of oath under no payoff commonality. The rationale is that if individuals followed their group members and lied out of pure conformity in part 2, one can expect them to continue to conform when they were alone in part 3. On the other hand, if some individuals also acted out of group pressure, one can expect less lying when these individuals were alone in part 3 because the action in the group setting was forced. While the pure conformity exists in both incentive structures, group pressure should exist only under payoff commonality.

The data indicate that the effect of pure conformity is very strong in the baseline conditions. Lying rates in an individual setting after group interactions are similar across BaseNoPC and BasePC (91% and 92%, respectively). More importantly, these lying rates do not differ from the rate in a group setting of their own treatment. This implies that pure conformity drives lying in a group setting in the baseline conditions, while the group pressure under payoff commonality may play a minor role without an oath. However, there seems to indeed be group pressure under the payoff commonality rule *with* an oath: the lying rate in OathPC decreases from 92% in part 2 to 82% in part 3. Regression analyses show that after controlling for pure conformity (as proxied by the number of liars in the group setting in part 2), individuals in the OathPC were about 12 percentage points less likely to lie later when they were alone, compared to all the other treatments. This implies that a high lying rate under oath in a group setting with payoff commonality was driven by pure conformity *and* group pressure, which crowds out the impact of the oath – although efficiency concerns may co-exist with group pressure.

Overall, the experimental findings presented in this chapter show that the role of an honesty oath, as a soft intervention to tackle dishonesty, can have its limits in a group context where individuals can communicate with their peers. The incentive structure that creates peer pressure on individuals to conform to the group can crowd out the desirable impact of an oath. The findings have implications concerning a careful design of oath interventions or professional codes of conduct in strict connection with the incentive structure in an organization.

The remainder of this paper is organized as follows. Section 2.2 briefly reviews the related literature. Section 2.3 outlines the experimental design and procedure. Section 2.4 describes the behavioral conjectures. Section 2.5 reports the results of the experiment. Section 2.6 discusses these findings and concludes.

## 2.2 Related Literature

This chapter is the first to examine the effect of an honesty oath against lying in a group context where individuals can communicate with each other. Previous studies on the impact of oath focused on an individual setting and have demonstrated that it induces truth-telling (Jacquemet et al., 2019) as it interferes with the justification process. When an individual has sworn to be honest, the moral cost of lying increases because lying under oath involves two moral transgressions (*i.e.*, telling a lie and breaking the oath). This makes lying under oath less attractive. Nonetheless, recent studies suggest that the effect of oath is heterogeneous and may be limited against small and partial lies as opposed to big and obvious ones (Heinicke et al., 2019; Jacquemet et al., 2021), and against lies that are mutually beneficial as opposed to self-serving ones (Jacquemet et al., 2021).<sup>1</sup> Whether the impact of oath persists in a group setting or not is an important research avenue because frauds and scandals often involve group of individuals, not an individual alone. More importantly, when companies implement oath or professional codes of conduct, it generally applies to all employees who may need to make ethical decisions with others. Therefore, the finding will shed light on the *limits* of an honesty oath in a group setting.

Previous studies on group dishonesty indicate that groups are more likely to lie than individuals<sup>2</sup> (*e.g.*, Weisel and Shalvi, 2015; Kocher et al., 2018), which can be explained by various reasons. First, studies on group decision-making have found that groups are more sophisticated than individuals and are more likely to make a self-interested choice in an economic game (see a review by Charness and Sutter, 2012). For instance, groups choose a lower number in a beauty contest game (Kocher and Sutter, 2005), send and return less money in a trust game (*e.g.*, Cox, 2002; Kugler et al., 2007), make and accept smaller offer in a ultimatum game (Bornstein and Yaniv, 1998). Related to lying, Sutter (2009) found in a Sender-Receiver Game that groups are more likely to use a sophisticated deception than individuals (*i.e.*, telling the truth believing that the message will not be followed). Second, the group setting allows for diffusion of responsibility (*e.g.*, Falk and Szech, 2013; Falk et al., 2020). This occurs when the action of an individual towards a group decision becomes less pivotal, leading to a lower perceived individual accountability. Thus, individuals can exploit a moral wiggle room by hiding behind others (*e.g.*,

<sup>1</sup>As pointed out in Chapter 1, there are notable exceptions to these findings. See Koessler et al. (2019) and Prima et al. (2020) for null findings in the lab and field. Also see Jacquemet et al. (2019) and Cagala et al. (2019) for sensitivity of the impact of an oath and framing of the language, and Cagala et al. (2019) and Cagala et al. (2021) for the backfiring effect of an oath.

<sup>2</sup>An exception is Castillo et al. (2022) who found that groups are not more dishonest than individuals when lying creates negative externality on a third-party.



Conrads et al., 2013; Weisel and Shalvi, 2015; Bauer et al., 2021). Third, groups may lie more than individuals because others can benefit from it (*see e.g.*, Wiltermuth, 2011; Gino et al., 2013). However, a recent study by Kocher et al. (2018) found that groups lie to the same extent irrespective of whether others can benefit from it or not, which may imply that other-regarding preferences play a minor role in a setting where individuals can communicate and update their beliefs about others. Lastly, studies on norm conformity suggest that individuals in a group tend to lie more because they can learn about the empirical norm of honesty (*i.e.*, what others do) by observing the action of their peers (*e.g.*, Gino et al., 2009; Diekmann et al., 2015; Bicchieri et al., 2022) or communicating with them (Kocher et al., 2018). That is, the dishonest shift in groups results from the erosion of the norm of honesty due to conformity.

By design, this chapter rules out sophistication and diffusion of responsibility and looks at the role of conformity and incentives. Studies on peer effects and lying showed that people tend to misbehave after being exposed to norm violations of others (*see* Keizer et al., 2008; Gino et al., 2009; Fosgaard et al., 2013; Innes and Mitra, 2013; Kroher and Wolbring, 2015), and they do so asymmetrically. That is, people become more dishonest after observing norm violations but *do not* become more honest after observing norm compliance.<sup>3</sup> For instance, Lefebvre et al. (2015) found in a lab experiment that social information about non-compliance increases tax evasion, while compliance does not reduce it. A recent study by Colzani et al. (2021) found the same asymmetry in a die rolling paradigm. Further, evidence showed that people prefer being paired with a peer of the same type to reduce the moral cost of misbehaving (*e.g.*, Gross et al., 2018; Charroin et al., 2021). This chapter contributes by examining the effect of peer conformity and lying under oath in a group setting.

In addition, this chapter investigates how the impact of an honesty oath may interact with two group incentives proposed by Kocher et al. (2018) – the payoff commonality rule. Under this rule, group members are required to coordinate on an action, and failure to do so results in a loss for the whole group. This payment scheme reflects an organizational structure and how the incentives of individuals in a group can be interconnected. Kocher et al. (2018) argued that comparing lying behavior across these two incentive schemes addresses “the relevance of the other-regarding concerns argument in group decision making that involves a trade-off between payoff maximization and norm compliance.” Given their findings of no difference in the lying rates, Kocher et al. (2018) interpreted that group communication allows for

---








<sup>3</sup>An exception is Innes and Mitra (2013) who documented symmetric conformity in the Sender-Receiver Game with samples in India, but not among those in the USA.

social learning about what others do, which then leads to behavioral adjustments towards dishonesty. In other words, individuals lie more in a group setting because they simply conform to others.

Nonetheless, a mechanism not addressed by [Kocher et al. \(2018\)](#) is the role of group pressure under payoff commonality. In addition to the pure conformity effect, an individual in a group that earns a common payoff may also lie out of peer pressure (*i.e.*, I lie because others want me to). This implies that a person may lie more in a group setting not only because he or she wants to do what the others are doing (pure conformity) but also because he or she succumbs to group pressure. Thus, while only conformity drives lying in a group setting under no payoff commonality, group pressure is an additional force that induces more lying when group members earn a common payoff. This chapter will take a closer look at these two mechanisms since they may matter more when an oath is introduced. The tension in the group created by the payoff commonality rule may become stronger under oath because the individual now faces a trade-off between morality strengthened by the oath and lying out of group pressure. On the other hand, without the payoff commonality rule, there is no pressure to break the oath for the group and everyone is free to make their decision – though one can still conform to his or her peers. Therefore, because of the group pressure, one may expect a larger impact of oath when group members do not earn a common payoff than when they do.

## 2.3 Experimental Design and Procedures

### Design

Participants played a variant of the observed cheating game used in [Kocher et al. \(2018\)](#) (see [Gneezy et al. \(2018\)](#) for the original task). They observed the video of the outcome of a die roll (*i.e.*, , , , ,  or ), which was randomly selected by the computer. Each face of the die was associated with different payoffs: each eye on the die represented one token, except for ‘’ which yielded zero tokens. Participants were asked to memorize the outcome and report it later. They earned the payoff depending on their report and could misreport the number to earn more tokens.

Similarly to [Kocher et al. \(2018\)](#), a mixed design was used. The within-subjects dimension consisted of three parts and varied in whether the decision was made individually or in a group context. In part 1, the game was played individually as described, without any social interactions. The decision in part 1 serves as a control for individual moral type. In part 2, participants were randomly matched

to form a group of three who all viewed the same video of the die outcome. Then, they could chat anonymously for five minutes via computers. The chat allowed for a free-form communication apart from revealing identity. Messages were shown to all the three group members. They could leave the chat at any point of time. The chat ended either after all group members had left the chat or the time had elapsed. After the chat, each group member reported the number individually as in part 1. The decision in part 2 is will be of main interest in this chapter. Finally, in part 3, participants played the game individually as in part 1. The decision in part 3 serves not only to replicate [Kocher et al. \(2018\)](#), but also will be used to explore if oath has any effect after individuals have interacted with their group members in part 2. No feedback was given until the very end of the session. Thus, participants learned about the decision of their group members at the end of the session.

There were four (2x2) between-subjects treatments varying in the group incentives and the existence of oath. The between-subjects dimension was introduced at the beginning of part 2 of the experiment. Thus, part 1 remained the same across treatments. In **BaseNoPC**, payoffs of the group members were independent. Participants were informed that the number they report would not affect the payoff of their group members, and vice versa. In **BasePC**, group members faced payoff commonality. They were informed that to earn the payoff associated with the number, all the three members must report the same number. Otherwise, they would earn nothing. BaseNoPC and BasePC serve as the baseline conditions where individuals lie in a group setting under two different incentive structures. The two treatments, **OathNoPC** and **OathPC**, followed the payoff commonality rule of their respective baseline, *except* that at the beginning of part 2 (before receiving the instructions), all participants swore an honesty oath in which they committed to be honest and would tell the truth for the rest of the experiment, and it was common knowledge.<sup>4</sup>

The oath-taking procedure differs from how it was typically implemented in the literature in two major aspects. First, oath-taking was compulsory for all participants, while many previous studies made it voluntary. This is a design choice, as in reality individuals within the same organization or a profession do not really have the option of not signing an oath. For example, employees in the Dutch banking sector are required to swear bankers' oath before taking up the office. Making an oath compulsory also eliminates the uncertainty regarding whether the group members have signed it or not. Moreover, previous studies documented very low decline rates of signing the oath. Thus, the oath-taking was made compulsory and was made

---

<sup>4</sup>The exact wording of the oath, translated from French is “*I swear upon my honor that for the rest of the experiment, I will be honest and always tell the truth.*” Participants took the oath by retyping via computer.

common knowledge among all participants within the session. Second, the oath was introduced in the middle of the experimental session and not upon arriving at the lab. In particular, at the beginning of part 2 and before receiving the instructions, participants swore the honesty oath. There is a trade-off here. It was done to remain close to [Kocher et al. \(2018\)](#), while having the lying decision in part 1 as a control for individual's moral type before oath-taking. A drawback is the observed effect of an oath can be at the lower bound because they have been exposed to a similar task in part 1. If anything, in reality, people may have had experience with cheating opportunities before signing an oath.

Finally, at the end of the session, unannounced, participants answered three questions about their beliefs about the empirical norm of honesty for each part of the experiment in the session. More precisely, they had to guess the proportions of participants reporting each die number in each part. For each belief question, they were informed that they would receive 5 Euros minus the penalty (*i.e.*, 0.04 Euros per percentage point deviation from the true value). To avoid any loss in case of extreme deviations, the minimum payment was fixed at 0.50 Euros. The program randomly selected one of the three belief questions for an additional payment. Participants then answered standard socio-demographic questions and received feedback about their payoffs for each part and for the belief question chosen by the program.

## Procedures

The experiment was run in-person at GATE-Lab in Lyon, France, between November and December 2021. A total of 23 sessions were conducted: 5 sessions with 99 participants in BasePC, 6 sessions with 102 participants each for BaseNoPC, OathNoPC and OathPC.<sup>5</sup> All 405 participants (54.6% females) were recruited via HRoot ([Bock et al., 2014](#)), mainly from local engineering, business and medical schools.<sup>6</sup> The experiment was programmed using z-Tree ([Fischbacher, 2007](#)) and used the videos of the die outcomes, downloaded from the website of one of the

---

<sup>5</sup>The number of observations is based on a preliminary power calculation. Assuming a directional hypothesis of the effect of an oath on lying, the power is based on a One-tailed test. Assuming a Type-I error rate of 0.05, a power level of 0.8 and a medium to large effect size (Cohen's  $d = 0.65$ ), the required number of observations to uncover the hypothesized effect between the treatments and the baseline conditions using One-tailed Mann-Whitney U test is 32 observations per treatment. Given that one group comprises of three individuals, this implies 96 individual observations per treatment. The actual number of observations is slightly higher than the initial power calculation to account for instances where participants saw a video of '🎲' in part 2, which left no room to observe lying behavior.

<sup>6</sup>See Table 2.B.1 in the Appendix for the summary statistics of the socio-demographic characteristics.

authors of [Kocher et al. \(2018\)](#).<sup>7</sup> As in [Kocher et al. \(2018\)](#), the randomization of the videos of the die outcomes was conducted for the first two sessions and then used for the rest of the sessions to increase statistical power and simplify treatment comparisons.<sup>8</sup>

Upon arrival, participants randomly drew a ticket from an opaque bag which assigned them to computer terminals. A general instruction was given and read aloud at the beginning of the session. The remaining of the instructions were shown on computer screens before the beginning of each relevant part.<sup>9</sup> The average duration of the sessions was 50 minutes.

Participants' additional earnings consisted of their payoffs from one randomly selected part and one randomly selected belief question. Before receiving feedback about the payoff, a participant randomly drew a raffle to indicate the ID of the participant who would roll a six-sided die to determine which part was payoff relevant. After the die was rolled, the participant was asked to announce the die outcome and the relevant part to the session. The average payment was 17.32 Euros (SD = 2.71), including a fixed fee of 5 Euros.

## 2.4 Conjectures

This section formulates behavioral conjectures regarding the impact of oath on lying in a group setting depending on the incentive structure.<sup>10</sup>

In part 2 of the experiment, participants formed a group of three and observed the same video of a die outcome. Before reporting the die number individually, they communicated anonymously via a computerized chat for five minutes. The baseline conditions differ in the incentive structure. In BaseNoPC, there was no payoff commonality rule. The number a group member entered did not affect the payoff of the other group members, and vice versa. In BasePC, group members faced payoff commonality where they needed to enter the same number to receive the associated

---

<sup>7</sup>The author is thankful for personal correspondence with one of the authors who shared the original program.

<sup>8</sup>Despite this initial plan, due to different turn-up rate across sessions, the videos shown are not perfectly balanced. The distributions of the videos shown are not different across sessions, which implies that the results are free from any differences in the videos shown.

<sup>9</sup>The instructions are provided in Section 2.A of the Appendix.

<sup>10</sup>The design and conjectures were pre-registered at AsPredicted (#79968) prior to running the experiment. A clerical error was made by the author where the belief elicitation stage was not mentioned in the pre-registration, despite having it in the design. The analysis of belief elicitations is reported along with other exploratory analyses (*i.e.*, group coordination and decision times under oath) in Section 2.E of the Appendix.

payoff. If at least one of them deviated, all members earned zero tokens. [Kocher et al. \(2018\)](#) found that groups cheated to a similar extent irrespective of the payoff commonality. Since BaseNoPC and BasePC closely follow [Kocher et al. \(2018\)](#)'s GroupNoPC and GroupPC treatments, a similar pattern of lying under these two incentive structures is expected.<sup>11</sup>

Evidence from the recent literature on oath have shown that an intervention by which individuals commit themselves to be honest is an effective instrument in reducing lies ([Jacquemet et al., 2019](#)) because it increases the moral cost of lying. In contrast with the literature, this chapter considers the case of lying in a group context. A group setting may weaken the impact of oath if individuals exchange justifications to break the oath. However, introducing an oath makes the honesty norm salient, which may shift the content of group communication towards an ethical one. Whether or not the impact of oath survives in a group context where group members can communicate with each other is an empirical question.

The first conjecture is formulated as follows:

**Conjecture 1:** *For a given payoff rule, an individual oath-taking reduces lying in a group setting. This should hold if the honesty pledge interferes with the justification process to lie by increasing the moral cost of lying.*

Extending from [Kocher et al. \(2018\)](#), the incentive structure may play an important role in lying in a group context when individuals are under oath. The payoff commonality rule requires all the group members to coordinate on their reports, which may create a pressure to conform to ‘what the group wants’, and this can be either in favor or against honesty. There is the honesty norm made salient by the oath, on the one hand. However, it can be impeded by other competing norms preferred by the group. For instance, the group may want to maximize gains for efficiency, uphold cooperation or even behave altruistically for one another. This is likely since lying under payoff commonality can be viewed as pareto-efficient, assuming that the majority of the group prefers to violate the norm of honesty and thus the oath. Thus, under payoff commonality, individuals may lie out of peer pressure in addition to the pure conformity.

---

<sup>11</sup>[Kocher et al. \(2018\)](#) elicited beliefs about the fraction of liars in a reference experiment ([Fischbacher and Föllmi-Heusi, 2013](#)) before part 2 and after part 3. They were interested in how group interactions shifted the beliefs about the empirical honesty in the reference experiment. As this chapter aims to test the impact of oath in a group setting, these elicitations of [Kocher et al. \(2018\)](#) are not in the design to avoid any confounding effect with the oath.

Such a pressure plays a smaller role when the incentive structure does not impose payoff commonality. Since the decision of an individual group member does not impact the other members and vice versa, this allows each member to follow the norm of honesty and thereby respect the oath irrespective of whatever the others choose to do – though one can still conform to others. In other words, the honesty norm is not at competition with the other norms of the group. Therefore, the impact of oath against lying in group is expected to be smaller when group members earn a common payoff compared to when they do not because of the group pressure.

The second conjecture is formulated as follows:

**Conjecture 2:** *The impact of an oath intervention against lying is larger when group members do not face payoff commonality than when they do. This should hold if the payoff commonality rule creates social pressure on group members to conform to the group.*

## 2.5 Results

This section is organized as follows. Section 2.5.1 provides an evidence that oath impacts lying in a group setting but only when group members do not earn a common payoff. Section 2.5.2 shows that payoff commonality rule creates the pressure on group members to lie. Section 2.5.3 analyzes the chat data and explores how oath affects the content of communication.

### 2.5.1 The Impact of Oath on Lying in Group

This section focuses on the decision to lie in a group context during part 2 of the experiment. A two-sided test is used in general except when testing the impact of oath. As conjectured that oath would reduce lying, there is a clear assumption regarding the direction of the effect and thus a one-sided test is applied. Tests are performed after excluding instances where ‘☒’ was shown, up to the part being analyzed (*i.e.*, analysis of part 2 excludes observations where ‘☒’ was shown in part 1 and 2).<sup>12</sup>

The fractions of liars in part 2 are very high in the baseline conditions: 91.4% lied in BaseNoPC and 94.9% in BasePC ( $p = 0.535$ , Fisher’s exact test), which are

---

<sup>12</sup>Note that the use of Fisher’s exact test deviates from the pre-registered analysis plan. Given that liars are identifiable at the individual level and the data in all treatments are in a group setting, Fisher’s exact test is more appropriate than Mann Whitney U test.

significantly higher than the lying rates in part 1: 57.3% in BaseNoPC and 56.9% in BasePC (both  $p < 0.0001$ , signrank tests). These lying rates in a group setting are similar to those observed in [Kocher et al. \(2018\)](#) (86.3% in GroupNoPC and 89.7% in GroupPC). Thus, the result in the baseline conditions replicate the findings of [Kocher et al. \(2018\)](#): a group setting shifts dishonesty and groups lie to a similar extent irrespective of the payoff commonality.

Introducing oath reduces lying in groups, but to a different extent depending on the payoff commonality rule. When group members do not earn a common payoff, oath reduces the fraction of liars (from 91.4% in BaseNoPC to 76.5% in OathNoPC,  $p = 0.008$ , one-sided Fisher’s exact test). The reduction is, however, *smaller* when group members earn a common payoff (from 94.9% in BasePC to 86.2% in OathPC  $p = 0.053$ , one-sided Fisher’s exact test). The difference-in-difference in lying in part 2 and part 1 is significant between BaseNoPC and OathNoPC ( $p = 0.007$ ) but not between BasePC and OathPC ( $p = 0.185$ , Fisher’s exact tests). It is noteworthy that despite the reduction in the fractions of liars in OathNoPC and OathPC, they are still quite high when compared to previous studies which examined the effect at an individual level.<sup>13</sup>

Figure 2.1 displays the distribution of reported points across treatments. Oath reduces the average reported numbers when there is no payoff commonality (4.9 in BaseNoPC vs. 4.4 in OathNoPC,  $p = 0.009$ , ranksum test). The reduction in the average reported numbers is only weakly significant when group members face the payoff commonality (4.9 in BasePC vs. 4.8 in OathPC,  $p = 0.076$ , ranksum test).<sup>14</sup> The impact of oath translates into monetary terms as follows. While the amounts overclaimed in BaseNoPC and BasePC were on average about 5.5 and 5.8 Euros, respectively ( $p = 0.481$ , ranksum test), oath reduces overclaiming to about 4.4 Euros in OathNoPC (vs. BaseNoPC,  $p = 0.039$ ) and 5.3 Euros in OathPC (vs. BasePC,  $p = 0.365$ ). This is consistent with the larger impact of oath under no payoff commonality. In monetary terms, oath reduces the experimenter’s loss from cheating but only when group members do not earn a common payoff.

To confirm the above findings, Table 2.1 reports the average marginal effects from Probit regressions where the dependent variable is a binary indicator whether

<sup>13</sup>The post-hoc effect size of the oath under no payoff commonality is 0.41. As a comparison point, [Cagala et al. \(2019\)](#) examined the effect of oath on lying using the observed cheating game in an individual setting. Their effect size is 0.39. Given both studies used an individual payment scheme, it implies that although group interactions shifted dishonesty upward, the effect size of the oath remains fairly similar.

<sup>14</sup>Using ranksum tests, all pairwise comparisons of the die numbers observed in part 2 across treatments are not significant. The lowest  $p$ -value is 0.526. Thus, the treatment comparisons cannot be explained by differences in the die outcomes shown.



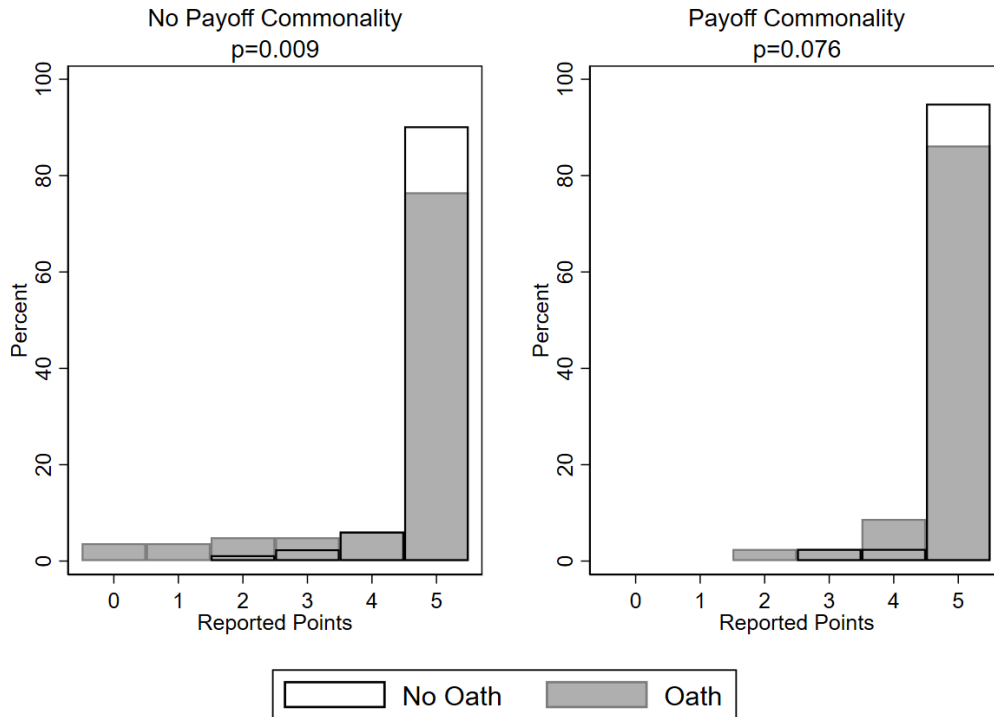


Figure 2.1: Distribution of Reported Points, by Treatment.

*Notes:* This figure displays the distributions of reported points by treatment. Left Panel reports BaseNoPC (White) and OathNoPC (Grey). Right Panel reports BasePC (White) and OathPC (Grey). Instances where ‘☒’ was observed are excluded.  $p$ -values reported using ranksum tests

the participant lied or not in part 2 (coded one for liar, zero otherwise). Model 1 indicates that participants in OathNoPC are about 15 percentage points less likely to lie. Model 2 suggests that this is robust after controlling for the points observed in part 2, individual characteristics (age, gender and self-reported risk attitudes) and moral type (with reference to part 1). In contrast, the impact of oath under payoff commonality is marginal: participants in OathPC are 5-6 percentage points less likely to lie, but the effect is far from being significant.

These analyzes support the following results:

**Result 1.** Oath reduces lying in a group setting, but *only* when group members do not earn a common payoff. If anything, the impact of oath when group members earn a common payoff is marginal. Thus, Conjecture 1 is partially supported.

**Result 2.** The impact of oath against lying in group is larger when group members do not earn a common payoff than when they do. Thus, Conjecture 2 is supported.

Table 2.1: Lying under Oath in a Group Setting

Dep. Var.	(1)		(2)	
Lying in Part 2	ME	St.Err.	ME	St.Err.
BaseNoPC	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
BasePC	0.035	(0.057)	0.021	(0.047)
OathNoPC	-0.149**	(0.070)	-0.159**	(0.062)
OathPC	-0.052	(0.074)	-0.064	(0.064)
Points Observed in Part 2	-	-	-0.052***	(0.017)
Age	-	-	-0.007	(0.004)
Male	-	-	-0.022	(0.034)
SOEP	-	-	0.006	(0.008)
Honest Reporting in Part 1	-	-	-0.119***	(0.036)
<i>N</i>	322		322	
Cluster	127		127	
Pseudo R-Squared	0.055		0.196	

*Notes:* This table reports the average marginal effects from Probit regressions. The dependent variable is a binary indicator whether the participant lied or not in part 2 (coded one for liar, zero otherwise). Instances where ‘☒’ was observed in part 1 and 2 were excluded. Model 1 includes treatment dummies in the independent variables (BaseNoPC as a reference category). Model 2 further controls for the die number observed in part 2 (reverse-coded 0 to 5), age (in years), gender (coded one for male, zero otherwise), self-reported risk attitudes (ranging from 0 to 10) and a dummy indicating if the participant was honest in part 1 or not (coded one for honest, zero otherwise). Robust standard errors clustered at the group level in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## 2.5.2 Group Pressure under Payoff Commonality

The previous section has shown that the impact of oath against lying in group is larger under no payoff commonality. This section provides an evidence behind this finding by looking at how individuals behaved in the group setting in part 2 and after group interactions in part 3. The analyses reported in this section exclude all instances where ‘☒’ was shown in all the three parts, leaving the remaining of 262 individual observations who had opportunities to lie in all parts.

In part 2, after observing the same video of the die outcome, participants could communicate with their group members via a computer chat for five minutes. They then reported the number of the die outcome individually, which determined their payoff, depending on the payment scheme of the assigned treatment. Group members in BasePC and OathPC faced the payoff commonality rule, which required them to report the same number to earn the associated payoff. The rule creates pressure on individuals to conform to what the group wants. Therefore, a person might lie out of *peer pressure* from the group members under payoff commonality at the cost of neglecting what he or she wants to do as an individual. Besides acting out of pressure, individuals may simply comply to others in their group out of *pure conformity* (*i.e.*, I do what others are doing). Thus, these two effects could drive people to lie in a group setting under the payoff commonality rule. On the other hand, the group pressure is absent in BaseNoPC and OathNoPC because group members independently received the payoff depending on the number they entered. Therefore, without the payoff commonality rule, individuals were under no pressure to report a certain number, but could still comply to others out of pure conformity after being exposed to the group chat.

If the payoff commonality rule indeed created the pressure to conform to what the group wanted in part 2, an individual who was under pressure should be less likely to lie when he or she was alone in part 3 because it was not what he or she wanted to do. On the other hand, if an individual lied out of pure conformity, he or she can be expected to continue to behave in a similar manner in part 3 because the action was not forced. If this reasoning holds, the fraction of liars in an individual setting after group interaction is expected to be *lower* when group members earn a common payoff compared to when they do not. The effect may be present in the baseline condition, and should be enhanced by the oath, assuming that introducing the oath made the honesty norm more salient.

The fraction of liars in an individual setting after group interactions is 91% and 92% in BaseNoPC and BasePC respectively ( $p = 1.000$ , Fisher’s exact test). These lying rates do not differ significantly from the rate in part 2 of their own treatment

(91% in BaseNoPC and 98% in BasePC, respectively  $p = 1.000$  and  $p = 0.218$ , signrank tests). This pattern of lying implies that pure conformity runs strong in the baseline conditions, while the payoff commonality rule might not have created pressure on individuals, resulting in similar rates of lying, both across treatments (in part 3) and within treatments (part 2 to 3).<sup>15</sup>

In OathNoPC, lying behavior after group interactions seems to show a sign of conformity. The fraction of liars in part 2 and 3 exhibits an increasing pattern – it rises from 76% in part 2 to 85% in part 3, but the difference is not significant ( $p = 0.1094$ , signrank test). This could imply that people might have conformed to others after being exposed to the justifications and norm violations of the group members. On the other hand, the fraction of liars in part 2 and 3 in OathPC has a decreasing pattern – it falls from 92% in part 2 to 82% in part 3. This pattern is in line with the pressure explanation that people in the group that faced payoff commonality might conform to the group out of pressure who then became less likely to lie later. The difference is, however, weak ( $p = 0.065$ , signrank test). This could be because certain individuals indeed succumbed to the pressure in part 2, while others might have complied out of pure conformity. These two effects then worked in the opposite direction in part 3. Comparing the treatments with their respective baseline, the fraction of liars in part 3 in OathPC is significantly lower than that of BasePC ( $p = 0.058$ ), while it is not significantly lower in OathNoPC than that in BaseNoPC ( $p = 0.212$ , one-sided Fisher's exact).<sup>16</sup>

To confirm these findings, Table 2.2 reports the average marginal effects from Probit regressions where the dependent variable is a binary indicator whether the participant lied or not in part 3 (coded one for liar, zero otherwise). Model 1 includes the treatment dummies while model 2 further controls for the points observed in part 3, individual characteristics and moral type (with reference to part 1) and the number of liars (self included) in the group in part 2, as a proxy of pure conformity. Results from model 2 suggest that participants in OathPC were about 12 percentage

<sup>15</sup>This is in line with post-questionnaire where participants were asked to rate on a likert scale from 1 to 7 regarding the level of pressure they felt during their decision making in part 2 (1 being no pressure to 7 being very high pressure). The mean response was 1.6 and 1.5 in BaseNoPC and BasePC, respectively ( $p = 0.267$ , ranksum test). On average, participants in OathNoPC reported that they did not feel more pressured than those in BaseNoPC (1.9 vs. 1.6,  $p = 0.495$ , ranksum test) while those in OathPC reported they felt more pressure than those in BasePC (2.0 vs. 1.5,  $p = 0.018$ , ranksum test).

<sup>16</sup>Interesting to note is that the lying rate in part 3 is significantly higher than that in part 1 for all treatments ( $p < 0.0001$  in both BaseNoPC and BasePC;  $p = 0.001$  in OathNoPC, and  $p = 0.002$  in OathPC, signrank tests). This implies that effect of the oath disappeared after group interactions.

Table 2.2: Individual Lying in Part 3: Pressure vs. Conformity

Dep. Var.	(1)		(2)	
	ME	St.Err.	ME	St.Err.
Lying in Part 3				
BaseNoPC	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
BasePC	0.014	(0.055)	-0.021	(0.040)
OathNoPC	-0.061	(0.059)	-0.011	(0.041)
OathPC	-0.094	(0.063)	-0.119***	(0.045)
Points Observed in Part 3	-	-	-0.043***	(0.012)
Age	-	-	-0.007	(0.005)
Male	-	-	-0.005	(0.035)
SOEP	-	-	-0.003	(0.007)
Honest Reporting in Part 1	-	-	-0.197***	(0.035)
Nb. of Liars in Part 2	-	-	0.083***	(0.016)
<i>N</i>	262		262	
Cluster	112		112	
Pseudo R-Squared	0.023		0.430	

*Notes:* This table reports the average marginal effects from Probit regressions. The dependent variable is a binary indicator whether the participant lied or not in part 3 (coded one for liar, zero otherwise). Instances where ‘☒’ was observed in part 1 and 2 were excluded. Model 1 includes treatment dummies in the independent variables (BaseNoPC as a reference category). Model 2 further controls for the die number observed in part 3 (reverse-coded 0 to 5), age (in years), gender (coded one for male, zero otherwise), self-reported risk attitudes (ranging from 0 to 10), a dummy indicating if the participant was honest in part 1 or not (coded one for honest, zero otherwise), and the number of liars in the group in part 2 (self included). Robust standard errors clustered at the group level in parentheses. \*\*\*  $p < 0.01$

points less likely to lie in part 3 compared to all the other treatments.<sup>17</sup> On the other hand, those in OathNoPC did not lie differently from those in BaseNoPC and BasePC. Interestingly, the marginal effect of the number of liars in the group in part 2 is significant and positive: an additional liar in a group in part 2 increases the likelihood that the individual will lie in part 3 by 8 percentage points, which is consistent with the pure conformity effect. This then implies the existence of the two effects in OathPC: individuals were less likely to lie in part 3 because they were under group pressure in part 2, while there is also the effect of pure conformity.

In all, individual lying behaviors in part 3 are in line with the pure conformity and pressure effects. Pure conformity effect exists in all treatments, suggesting that individuals lied more in a group setting out of conformity and continued to do so when they were alone. On the other hand, the payoff commonality rule did not seem

<sup>17</sup>Wald tests; OathPC vs. BasePC,  $p = 0.036$ ; OathPC vs. OathNoPC,  $p = 0.002$ .

to create pressure in the baseline, but did so when oath was introduced. Therefore, the group pressure, *in addition to* pure conformity, pushed individuals to lie under oath when group faced payoff commonality, crowding out the impact of an oath.

These analyses support the following result:

**Result 3.** In addition to lying in group out of pure conformity, the payoff commonality rule puts pressure on the individual under oath to conform to the group. This results in the crowding out of the impact of oath under payoff commonality.

### 2.5.3 Chat Analysis

To understand better how group communication affects the decision to lie in part 2, and whether or not oath changes the way group members communicate, messages sent during chat were categorized by two independent research assistants who were blind to the objective of the experiment. More specifically, the two coders classified the messages into pre-defined categories based on the argument explicitly made (*i.e.*, whether the argument was made for or against honesty *and* related to money, honesty, insecurity, rules, others, consequences or oath). These categories are taken from Kocher et al. (2018), except the ones about the oath. A message may belong to more than one category. Thus, for each group chat, the coders counted the number of times messages with explicit arguments were sent for each category.<sup>18</sup>

This section proceeds by presenting the descriptive statistics of group chat then explores the content of the chat by the types of argument (honest vs. dishonest) used by the group members. Finally, the analysis examines the themes of arguments (*i.e.*, whether the argument was made related to money and morality).

#### Descriptive Statistics of Chat

The average chat duration is 180 seconds ( $SD = 89.9$ ), during which group members exchanged about 19 messages ( $SD = 10.4$ ).<sup>19</sup> Table 2.3 shows the average chat duration and volume of messages by treatment. Interestingly, oath does not impact how long group members communicate nor the volume of messages. As oath does not make a difference to chat duration and volume of messages, pooling the data

---

<sup>18</sup>As coders might interpret the messages differently, their counts were averaged, which to some extent minimize the bias in case of extreme opinions. The code book used by the two coders is reported in Section 2.D in the Appendix.

<sup>19</sup>One subject in BaseNoPC did not contribute to the chat. Chat data missing for one group in OathNoPC as none participated in the chat stage. Two groups per treatment observed ‘☒’ and are excluded.

together based on the incentive structure reveals that the payoff commonality rule reduces both the duration and volume ( $p = 0.0006$  and  $p = 0.031$ , respectively ranksum tests). Thus, it is the group incentive, and not the oath, that determines the length and frequency of group communication.

Table 2.3: Chat Duration and Volume of Messages

	(1)	(2)	(3)	(4)	(1-2)	(1-3)	(2-4)	(3-4)
	BaseNoPC	BasePC	OathNoPC	OathPC	<i>p</i> -values			
Duration (in sec.)	204.25 (91.40)	158 (89.62)	212.87 (82.24)	145 (80.51)	0.046**	0.926	0.559	0.003***
Volume of messages	22.86 (11.70)	17.77 (11.74)	17.97 (9.29)	16.19 (7.49)	0.024**	0.114	0.899	0.474
<i>N</i>	32	31	31	32				

*Notes:* This table presents average chat duration (in seconds) and the volume of messages by treatment. Standard deviations in parentheses. Figures are reported at the group level. *p*-values reported using ranksum tests. \*\*  $p < 0.05$  \*\*\*  $p < 0.01$

Looking at the content of communication, the majority of groups (66.7%) used dishonest arguments at least once, while honest arguments were made at least once in only 33.3% of the groups ( $p < 0.0001$ , Fisher's exact test).<sup>20</sup> Defining participant's moral type based on their lying decision in part 1, the first proposition about which number to report in part 2 was made more by dishonest individuals than honest individuals (67% vs. 33%,  $p = 0.050$ , Fisher's exact test).<sup>21</sup> Overall, the average share of dishonest arguments (54.0%) is higher than honest arguments (16.0%,  $p < 0.0001$ , within-subjects comparison at the group level using signrank test).<sup>22</sup> This pattern holds for all treatments. However, it is noteworthy that the difference between the shares of dishonest and honest arguments is the least in OathPC ( $p = 0.095$ , signrank test), while the gap between the shares are significantly larger in all the other treatments (all  $p < 0.001$ , signrank tests). Comparing across treatments, oath does not reduce the share of dishonest arguments, ranging from 55% to 61%.<sup>23</sup> Nonetheless, the share of honest arguments is higher under oath: compared to the pooled baseline, the share of honest arguments are higher in the pooled oath treatments (20% vs. 12%,  $p = 0.054$ , ranksum test).

In sum, the payoff commonality rule, not the oath, affects the duration of the chat

<sup>20</sup>See Table 2.B.2 in the Appendix for the number of groups by types of arguments in which honest and/or dishonest argument is mentioned at least once.

<sup>21</sup>Using Fisher's exact tests, there are no differences in the proportion of dishonest or honest proposers across treatments.

<sup>22</sup>Shares of honest and dishonest messages are calculated at the group level (*i.e.*, the number of times messages encouraging honesty were sent divided by the total number of times messages encouraging honesty and dishonesty were sent in a group). This makes the sum of shares of honest and dishonest messages equal 100% at the group level, but not at the treatment level or the global level.

<sup>23</sup>See Table 2.B.3 in the Appendix for shares of dishonest and honest messages by treatment.

and the volume of the messages. Group members used dishonest arguments more often than honest ones. Introducing oath increases the share of messages encouraging honesty, but does not reduce the share of messages encouraging dishonesty.

### Themes of Arguments

This section explores the chat data based on the explicit arguments made in the messages that were identified by the two independent coders to be related to different categories *and* either in favor of honesty or dishonesty. The categories were taken from Kocher et al. (2018), in addition to two categories relating to an oath.<sup>24</sup> Messages can be used to encourage (dis)honesty in relation to **Money** (*i.e.*, ‘Reporting 4 is not so bad, it gives us 8 Euros’, ‘We should report 5 to maximize our gains’), **Morality** (*i.e.*, ‘We should be honest’, ‘Respect the oath’, ‘Lying is ok’), **Instructions** (*i.e.*, ‘We should stick to the rule/ what the instruction says’, ‘It is not mentioned that we cannot lie’) or **Empirical Norms** (*i.e.*, ‘I think other people are honest’, ‘Many people will lie’). The analysis focuses on the arguments related to Money and Morality, as they are the major themes of group discussion.<sup>25</sup>

Figure 2.2 displays the frequencies of arguments related to Money and Morality, where the upper panel reports for dishonest use, and the lower panel reports for honest use.<sup>26</sup> Participants frequently used ‘money’ to encourage group members to lie irrespective of the treatments.<sup>27</sup> Arguments related to money in favor of honesty are rare (except OathPC where the frequencies increase but are not different from BasePC ( $p = 0.532$ , Fisher’s exact test)). This means that money related arguments used to encourage lying remain unimpeded by the oath.

On the other hand, participants under oath frequently made morality related arguments in favor of truth-telling than those in the baseline conditions (BaseNoPC vs. OathNoPC:  $p = 0.076$ ; BasePC vs. OathPC:  $p = 0.057$ ; pooled baselines vs. pooled oath treatments:  $p = 0.008$ , Fisher’s exact tests). As for the dishonest use, oath does not impact the frequencies of immoral arguments to encourage lying under no payoff commonality (BaseNoPC vs. OathNoPC:  $p = 0.868$ ), but these immoral arguments *increase* under oath when group members face payoff commonality

<sup>24</sup>Given a rather large number of categories to begin with, they are afterwards combined for ease of analysis. The category ‘Morality’ contains both arguments about ‘Honesty’ and ‘Oath’. The category ‘Instructions’ contains both arguments about ‘Rules’ and ‘Insecurity’.

<sup>25</sup>Shares of arguments in favor of dishonesty: 58% are related to Money, followed by Morality (24%), Instructions (14%) and Empirical Norms (4%). Shares of arguments in favor of honesty: 59% are related to Morality, followed by Money (20%), Instructions (20%) and Empirical Norms (1%).

<sup>26</sup>Please see Figure 2.C.1 in the Appendix for the frequencies of all types of arguments.

<sup>27</sup>No significant differences across treatments using Fisher’s exact tests. Comparing pooled baselines and pooled oath treatments yields  $p = 0.393$ .



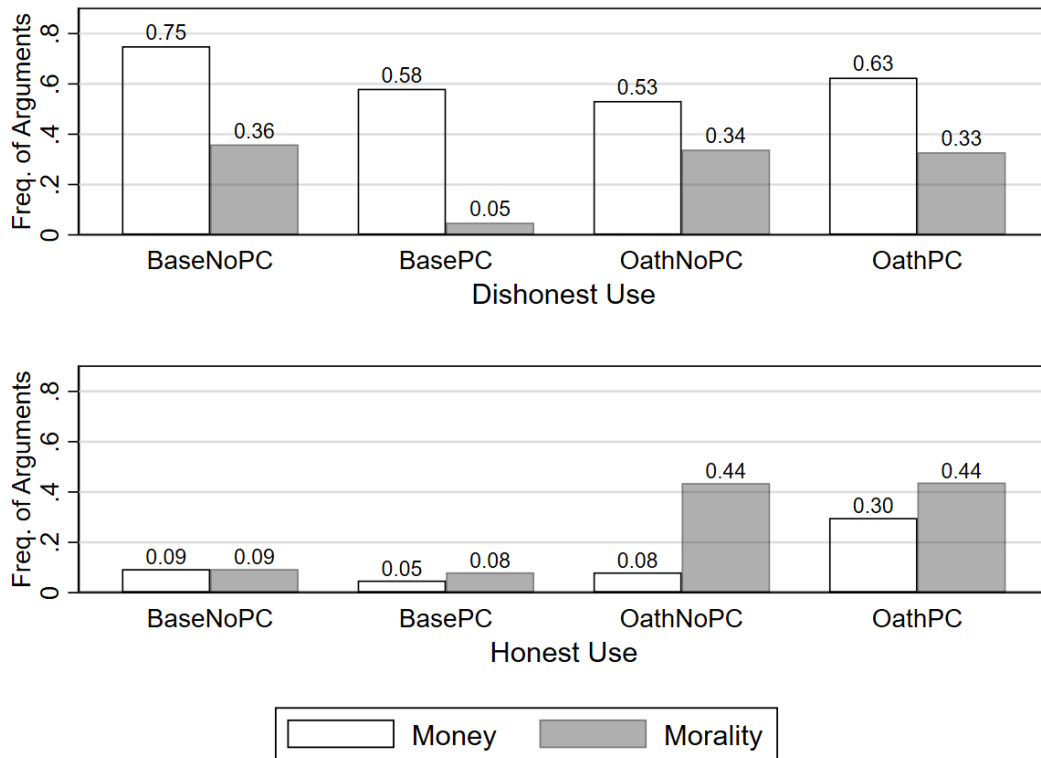


Figure 2.2: Frequencies of Arguments Related to Money and Morality

*Notes:* This figure displays the frequencies of arguments related to Money and Morality. Dishonest use (Upper Panel) Honest Use (Lower Panel).

(BasePC vs. OathPC:  $p = 0.022$ , Fisher's exact tests). The increase in morality related arguments to encourage lying under oath with payoff commonality is consistent with the group pressure effect. In other words, the payoff commonality rule may push group members to convince the others to lie under oath by justifying that lying is ok or downplaying the oath.

In sum, the chat data shows that group members favor lying to maximize their gains (as documented in [Kocher et al., 2018](#)) and this still remains a core discussion even when they are under oath. The honesty norm becomes more salient when oath is introduced, which is reflected in an increase in morality related arguments favoring honesty. However, as the payoff commonality rule exerts group pressure, participants who favor dishonesty might 'fight back' by downplaying the value of honesty or that of the oath through justifications. This is reflected in an increase in the use of immoral arguments in OathPC compared to BasePC.

Finally, to explore how each type of arguments affects the lying decision in the group setting, Table 2.4 reports the average marginal effects from Probit regressions where the dependent variable is a binary indicator whether the participant lied or not in part 2 (coded one for liar, zero otherwise). Model 1 includes in the

Table 2.4: Impact of Arguments on Lying in Part 2

Dep. Var.	(1)		(2)	
Lying in Part 2	ME	St.Err.	ME	St.Err.
BaseNoPC	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
BasePC	0.021	(0.047)	0.054	(0.052)
OathNoPC	-0.159**	(0.062)	-0.072	(0.062)
OathPC	-0.064	(0.064)	0.028	(0.056)
Points Observed in Part 2	-0.052***	(0.017)	-0.026*	(0.014)
Age	-0.007	(0.004)	-0.001	(0.004)
Male	-0.022	(0.034)	-0.016	(0.026)
SOEP	0.006	(0.008)	0.002	(0.006)
Honest Reporting in Part 1	-0.119***	(0.036)	-0.096***	(0.030)
Money related (dishonest use)	-	-	0.106***	(0.036)
Money related (honest use)	-	-	-0.016	(0.043)
Morality related (dishonest use)	-	-	0.014	(0.033)
Morality related (honest use)	-	-	-0.141***	(0.034)
Instruction related (dishonest use)	-	-	0.077	(0.047)
Instruction related (honest use)	-	-	-0.062	(0.054)
<i>N</i>	322		320	
Cluster	127		126	
Pseudo R-Squared	0.196		0.395	

*Notes:* This table reports the marginal effects from Probit regressions. The dependent variable is a binary indicator whether the participant lied or not in part 2 (coded one for liar, zero otherwise). Instances where ‘☒’ was observed in part 1 and 2 were excluded. One group in OathNoPC (of which one member saw ‘☒’ in part 1) did not contribute to chat and is excluded. Model 1 includes treatment dummies in the independent variables (BaseNoPC as a reference category), the die number observed (reverse-coded 0 to 5), age (in years), gender (coded one for male, zero otherwise), self-reported risk attitudes (ranging from 0 to 10) and a dummy indicating if the participant was honest in part 1 or not (coded one for honest, zero otherwise). Model 2 further includes the types of arguments made by group members (in favor of (dis)honesty *and* related to Money, Morality, Instructions). The category ‘Empirical Norms’ is omitted due to collinearity. Robust standard errors clustered at the group level in parentheses. \*  $p < 0.10$ , \*\*\*  $p < 0.01$

independent variables treatment dummies and control variables (which is the same as Model 2 of Table 2.1). Model 2 includes the arguments used in the chat (the category ‘Empirical Norms’ is omitted due to collinearity). It can be seen that dishonest arguments referring to money increase the likelihood to lie, while morality related arguments favoring honesty decrease it. Note that since morality related arguments to encourage truth-telling increased in OathNoPC compared to BaseNoPC, controlling for these arguments explains most of the treatment effect.

These analyses support the following result:

**Result 4.** Oath does not impact the duration of group communication and the volume of messages. However, it changes the way people communicate, particularly related to morality. Morality related arguments to discourage lying increases under oath. However, groups under oath that face payoff commonality use more immoral arguments to encourage lying by downplaying morality. This is consistent with the group pressure under oath with payoff commonality.

## 2.6 Discussion and Conclusion

This chapter investigates the impact of individual oath-taking against lying in a group setting. An oath may encourage individuals to promote moral behavior with the others in their group. However, group interactions also allow members to exchange justifications to lie, which may eventually compromise the impact of oath. In addition, it explores whether the impact of oath depends on the incentive structure group members face. In particular, it examines the case of the payoff commonality rule proposed by [Kocher et al. \(2018\)](#). Under this rule, group members are required to coordinate on their reports, and in case of any deviation, the group earns nothing. In a setting where communication is possible, such an incentive scheme can create pressure, in addition to pure conformity, on individuals to conform to what the group wants. This group pressure is absent under no payoff commonality rule. Thus, it is possible that the impact of oath against lying is larger when group members do not earn a common payoff than when they do owing to the presence of group pressure.

Results from a laboratory experiment reveal that the impact of oath against lying extends to a group setting, but *only* when group members do not earn a common payoff. The impact of oath is marginal under the payoff commonality rule. Analysis of individual lying behavior after group interactions provides supporting evidence of a group pressure under oath *with* payoff commonality. The findings reveal that individuals lied more in a group setting because they conformed to what the others

were doing, and continued to do so when they were alone in part 3. This is consistent with the pure conformity effect and it exists in all treatments. On the other hand, group pressure from the payoff commonality did not seem to play a role in the baseline condition, while it did when the oath was introduced – individuals under oath in a group that faced payoff commonality were less likely to lie in part 3. This means that, in addition to pure conformity, group pressure pushed individuals to lie under oath in a group setting when they faced payoff commonality, thereby crowded out the impact of an oath.

Chat analysis reveals how the oath changes the course of group communication. Individuals primarily deploy arguments related to money to convince their group to lie, and this remains unimpeded by the oath. Nonetheless, introducing oath increases the frequency with which individuals use moral arguments to encourage honesty under both incentive schemes, but more immoral arguments are made under oath when group members face payoff commonality compared to its baseline without oath. Additional insights come from analyses of group coordination, beliefs about the empirical norms and decision times (see Section 2.E of the Appendix). Oath reduces coordination when groups do not face payoff commonality by moving from coordinating on dishonest reporting to remaining uncoordinated. However, when there is payoff commonality, oath has no impact on coordination rates. Belief about the empirical norms is consistent with the actual behavior. Participants expect oath to reduce group dishonesty and to lesser extent when group members earn a common payoff, and that group interactions would shift individual dishonesty upwards, while oath would counteract this shift. Lastly, individuals under oath spend significantly more time in deciding to lie or not, but only when there is no payoff commonality rule, suggesting that deliberation is more complex in this setting. Oath does not affect the decision time when group members earn a common payoff.

Concerning the mechanism of group pressure, the reason why individuals who lied in a group setting with payoff commonality were less likely to lie in an individual setting afterwards may stem from an image concern. If participants believe that they have lied *a lot* with the group, they may be more reluctant to lie again to dissociate themselves from bad apples and to restore their image with the experimenter (or even in their own eyes). These motives can indeed co-exist in parallel with group pressure. Nonetheless, they cannot be excluded nor do they change the interpretation of conformity out of peer pressure at the cost of negligence of individual honesty, which is then given attention afterwards. This behavior is also consistent with moral balancing (see [Monin and Miller, 2001](#); [Ploner and Regner, 2013](#); [Rahwan et al., 2018](#)) – individuals who misbehaved under pressure were less likely to lie afterwards. Finally, although [Kocher et al. \(2018\)](#) argued that other-

regarding preferences play a minor role under the two incentive schemes when oath does not exist (a finding replicated in this chapter), one may argue that these preferences may play a role when an oath is introduced. Although this motive cannot be entirely disentangled from the data, one can expect that payoff maximization should induce peer pressure on the individual group member to lie.

Another interesting point is that oath loses its power *after* group interactions as evident from a higher fraction of liars in part 3 compared to part 1 for all treatments. This implies that even though the oath achieved its desirable impact against lying in group (atleast under no payoff commonality), individuals had been exposed to the justifications of others during the group discussion, which could dampen the effect of oath on individual lying behavior. This is in contrast to the findings of [Peer and Feldman \(2021\)](#) where they found that the impact of oath is persistent in a repeated game setting of the individual lying task. This also largely holds for Chapter 1, which used the strategic lying task with negative externality. Nonetheless, the participants in those two experiments could not observe behaviors of others nor could they communicate with their peers, while participants in the experiment presented in this chapter had an opportunity to exchange justifications and observe what their peers think. This could be an interesting avenue to explore.

Overall, the findings of Chapter 2 suggest that while an oath intervention can be used to foster honesty in a group setting, its desirable impact can be limited when group members earn a common payoff due to peer pressure to conform to the group. One still needs to remain cautious about the external validity of these findings in a more complex environment like that of a real organizational setting. Nonetheless, they highlight how practitioners must pay close attention to the way an oath or a compulsory code of conduct is introduced within an organizational context as it must be designed in close connection with the incentive schemes. In addition, it suggests that oath is not a *be-all-end-all* policy to globally limit dishonesty in a collective setting. Thus, firms may benefit from optimizing this soft intervention, for example, by incorporating the social identity dimension such as proximity, which has been shown to halt norm erosion caused by peer conformity ([Bicchieri et al., 2022](#)), or implementing it in tandem with ‘hard’ interventions, such as internal auditing.

## Appendix

### 2.A Instructions (Translated from French)

The general instruction was given in print and read aloud at the beginning of the session. Then, the instruction for each part of the experiment was given via the computer screen. Instructions, both general and on-screen, and comprehension questions are kept as close as possible to those used in [Kocher et al. \(2018\)](#).

#### Instructions

Hello and welcome to this experiment in decision making. Please turn off your phone and put it away. You are not allowed to communicate with other participants during the experiment, unless invited to do so by the experimenters. You may be excluded from the session and gains. During this session, you can earn money. The amount you will earn depend on your decisions and the decisions of other participants in the session. Therefore, please read the instructions carefully.

#### Your earnings

During the session, we will not talk in Euros but in tokens. The conversion rate is as follows:

$$1 \text{ Token} = 2 \text{ Euros}$$

This session consists of three successive and independent parts. One of the three parts will be randomly selected to determine your earnings at the end of the session. At the end of the session, a randomly selected participant will roll a die. If the die displays  $\square$  or  $\boxtimes$ , your gains in the first part will determine your earnings for the session. If the die displays  $\boxdot$  or  $\boxminus$ , your gains in the second part will determine your earnings for the session. And if the die displays  $\boxplus$  or  $\boxtimes$ , your gains in the third part will determine your earnings for the session. Since all the parts have the same probability of being selected for payment, it is in your best interest to make your decision in each part as if that part were the one that counts towards your earnings.

The amount earned at the end of this session is the sum of your tokens in the selected part and your participation fee of 5 Euros. No one will know how much you have won, nor will you be informed of the earnings of the other participants.

You will be paid by transfer by the CNRS to your bank account. To do this, at the end of the session of the session we will ask for your IBAN. We will also ask you to to send us at the end of the session a bank statement (containing IBAN) in pdf format with the title “your first name-your last name” to the following email address: [gatelab\[at\]gate.cnrs.fr](mailto:gatelab@gate.cnrs.fr).

Given the administrative and banking delays, the transfer may take up to two weeks. We commit ourselves that your banking information will be used only for this transfer. This information will be kept separate from the files containing your decisions and the researchers who will process your decisions will not have access to your banking information.

**Anonymity of your decisions**

All your decisions and responses will be anonymous. We will never link your name with the data generated in the experiment. You will not know the identity of the other participants, either before or after the experiment. The other participants will not know your identity either.

At the beginning of each part, you will receive detailed instructions on your screen. If you have any questions about the instructions or during the experiment, please raise your hand or press the red button on the side of your desk. We will come and answer your questions immediately in private.

-END OF GENERAL INSTRUCTIONS-

-Instructions for Part 1-

*(Same for all treatments)*

**The die roll**

In part 1, you will see a video of a die roll, showing (1, 2, 3, 4, 5 or 6).

The video you will see is chosen randomly by the program and each of the 6 videos is chosen with the same probability. *(For session 3 and onward: These draws have been conducted before the start of the session.)*

**Your task**

Your task is to memorize the outcome of the die roll shown in the video and enter it later.

**Your gains for this task** The number you enter determines your gains in tokens as follows:

Number Entered	'1'	'2'	'3'	'4'	'5'	'6'
Tokens Gained	1	2	3	4	5	0

**Comprehension Questions**

1. What is your task in this part?
  - Enter the number displayed that you have memorized
  - Enter the number different from the one seen and memorized
  - Enter a random number
2. Suppose that you saw 3 and you enter “3”, how many tokens do you receive?
3. Suppose that you saw 3 and you enter “2”, how many tokens do you receive?
4. Suppose that you saw 3 and you enter “4”, how many tokens do you receive?

-Before Instructions for Part 2-

*(OathNoPC and OathPC Only)*

Before continuing the experiment, we ask all participants to take an oath.

With this oath, each participant swears that for the remainder of the experiment, he or she will be honest and will always tell the truth.

Please click “Next” to take the oath.

(Next Button)

You must now swear the following oath: “I swear upon my honor that for the rest of the experiment, I will be honest and that I will always tell the truth.”

Please write this oath in the box below, then click the ”Next” button.

(Next Button)

-Instructions for Part 2-

*(Same for all treatments, except sentence about the payoff rule)*

Part 2 of the experiment is similar to part 1, except that you will now decide in group. You will be randomly matched with two other participants to form a group of three.

### **The die roll**

In part 2, you will see a video of a die roll, showing (1, 2, 3, 4, 5 or 6).

The video you will see is chosen randomly by the program and each of the 6 videos is chosen with the same probability. *(For session 3 and onward: These draws have been conducted before the start of the session.)* All the three members in your group will see the same video chosen at random.

### **Your task**

Your task is to memorize the outcome of the die roll shown in the video and enter it later.

**Your gains for this task** The number you enter determines your gains in tokens as follows:

Number Entered	‘1’	‘2’	‘3’	‘4’	‘5’	‘6’
Tokens Gained	1	2	3	4	5	0



*(The sentence for BaseNoPC & OathNoPC)*

The numbers entered by the other two group members do not affect your gains in tokens in this part. Similarly, the number you enter does not affect the gains of the other members of your group.

*(The sentence for BasePC & OathPC)*

The numbers entered by the other two group members affect your gains in tokens in this part. If the numbers entered by all group members are not the same, all members of your group receive zero tokens in this part.

### Comprehension Questions

1. What is your task in this part?
  - Enter the number displayed that you have memorized
  - Enter the number different from the one seen and memorized
  - Enter a random number
2. Suppose that you saw  $\square$  and you enter “3”, how many tokens do you receive, if ...
  - the members in your group enter the same number?
  - the members in your group enter a different number?
3. Suppose that you saw  $\square$  and you enter “2”, how many tokens do you receive, if ...
  - the members in your group enter the same number?
  - the members in your group enter a different number?
4. Suppose that you saw  $\square$  and you enter “4”, how many tokens do you receive, if ...
  - the members in your group enter the same number?
  - the members in your group enter a different number?

-Chat Stage in Part 2-

*(Same for all treatments)*

You now have the opportunity to chat with the other two members of the group to clarify about the number that each group member must enter.

You have 5 minutes to exchange information. The group chat ends after 5 minutes or as soon as all the 3 group members have pressed the button “Leave chat”. If only 1 or 2 group members press the button, the chat will continue either until all of them press the button or until the time is up. If you have pressed the button “Leave Chat” but do not want to leave the chat, you can press the button “Back”. After the group chat, each member of the group enters the number of the die roll.

You are free to choose the content of the communication, but it is not allowed to mention any personal information (such as name, age, gender, school, field of study or any other identifying details, e.g., seat number). In addition, you are not allowed to agree to any side payment outside of the experiment. If you break any of these rules, you will be excluded from the experiment and will not receive any payment.

During the chat, each group member can send as many messages as they want to other group members. Each of your messages will automatically appear on the screen of the other two group members. It is not possible to send messages to only one person.

The chat screen looks like this

*(Screenshot of Chat Stage appears here)*

To write a message, click on the purple tab, type your message and press the enter key on your keyboard. Your message will appear in the gray box above. Other participants in your group will not see your message until you press “Enter”.

### **Comprehension Questions**

Suppose that you have pressed the button “Leave Chat”, when do you leave the chat stage?

- Immediately
- When you press the button “Back”
- When all the members of your group have pressed the button “Leave Chat” or when the time limit expires

-Instructions for Part 3-

*(Same as in part 1 and in all treatments)*

-Belief Elicitation Screens-

*(Same for all treatments)*

### **Question about part 1**

Please answer as accurately as possible.

In part 1, each participant saw his or her video of the die outcome and reported the number.

*(Actual distribution of the video of die outcomes displayed here)*

Please estimate the percentages of participants who entered each of the numbers in Part 1. If at the end of the session, this question is selected by the program, you will receive 5 euros. For each percentage point by which your answers differ from the correct values, your payout will be reduced

by 4 cents. The smallest possible amount for this question is 50 cents.

- What is the percentage of participants that entered the number “1”?
- What is the percentage of participants that entered the number “2”?
- What is the percentage of participants that entered the number “3”?
- What is the percentage of participants that entered the number “4”?
- What is the percentage of participants that entered the number “5”?
- What is the percentage of participants that entered the number “6”?

### **Question about part 2**

Please answer as accurately as possible.

In Part 2, each participant was assigned to a group of three, watched the same video of the die roll result, and reported the number after the group discussion.

*(Actual distribution of the video of die outcomes displayed here)*

Please estimate the percentages of participants who entered each of the numbers in Part 2. If at the end of the session, this question is selected by the program, you will receive 5 euros. For each percentage point by which your answers differ from the correct values, your payout will be reduced by 4 cents. The smallest possible amount for this question is 50 cents.

- What is the percentage of participants that entered the number “1”?
- What is the percentage of participants that entered the number “2”?
- What is the percentage of participants that entered the number “3”?
- What is the percentage of participants that entered the number “4”?
- What is the percentage of participants that entered the number “5”?
- What is the percentage of participants that entered the number “6”?

-Belief Elicitation of Part 3-

*(Same as the belief elicitation of part 1)*

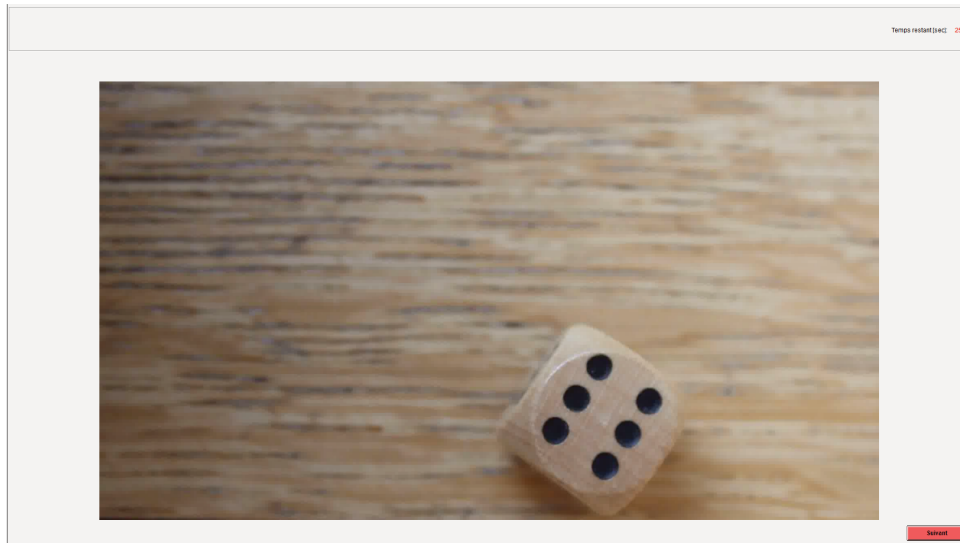


Figure 2.A.1: Sample Video of Die Outcome

*Notes:* This figure displays the computer screen in which a die outcome is shown through a video. The videos are downloadable from the website of one of the authors of [Kocher et al. \(2018\)](#).

Temps restant (sec) 115

Saisissez maintenant le résultat du lancer de dé dans la vidéo, puis appuyez sur "Suivant".

Le nombre vu:

Le nombre du dé que vous saisissez détermine votre gain en jetons comme suit :

Nombre saisi:	"1"	"2"	"3"	"4"	"5"	"6"
=	=	=	=	=	=	=
Jetons	1	2	3	4	5	0

Suivant

Figure 2.A.2: Decision Screen in Part 1

*Notes:* This figure displays the decision screen in part 1 of the experiment.

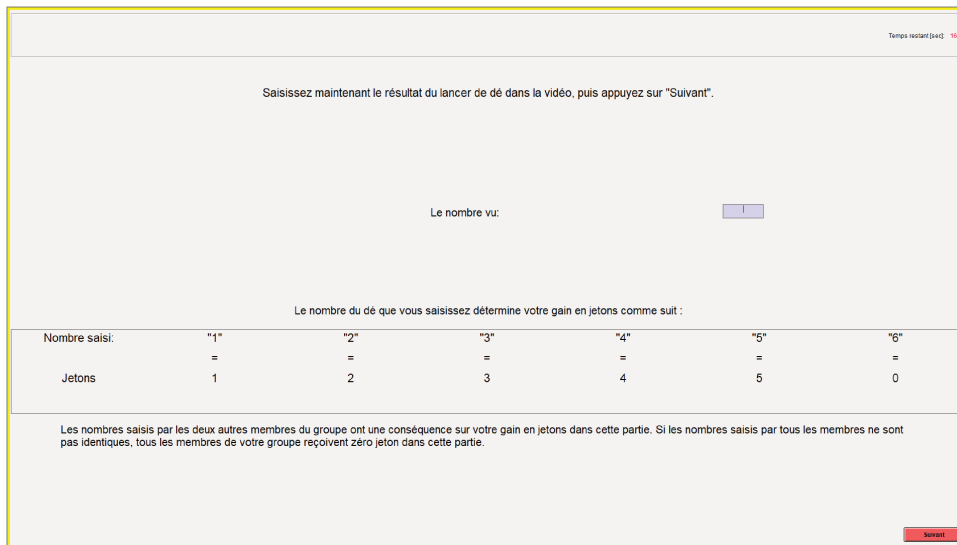


Figure 2.A.3: Decision Screen in Part 2

*Notes:* This figure displays the decision screen in part 2 of the experiment. The description below the payoff table varies according to the payoff commonality rule.

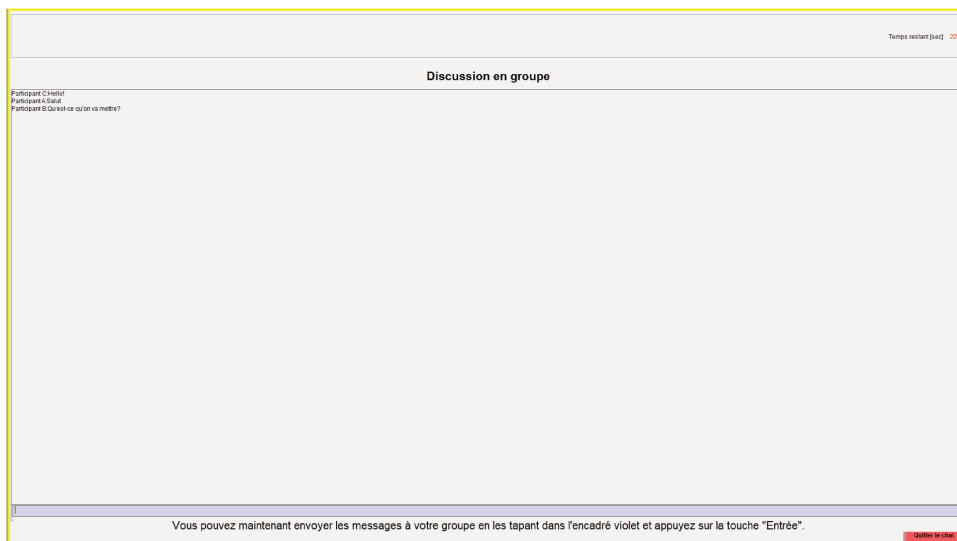


Figure 2.A.4: Sample Chat Stage in Part 2

*Notes:* This figure displays the chat stage in part 2 of the experiment. Participants can type the messages on the purple bar below. After pressing "Enter", the message are sent to the whole group, appearing in the grey box above.

## 2.B Appendix Tables

Table 2.B.1: Summary Statistics

	(1)		(2)		(3)		(4)		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)
	BaseNoPC		BasePC		OathNoPC		OathPC							
	Mean	SD	Mean	SD	Mean	SD	Mean	SD						
Age (Years)	21.64	4.09	21.23	1.96	22.08	3.10	20.85	1.73	0.796	0.013**	0.342	0.024**	0.219	0.001***
Male (%)	0.44	0.50	0.45	0.50	0.45	0.50	0.46	0.50	0.849	0.888	0.778	0.960	0.929	0.888
SOEP (0-10)	6.26	2.39	6.26	2.12	6.13	2.43	5.57	2.30	0.841	0.843	0.027**	0.945	0.037**	0.051**
Business School (%)	0.54	0.50	0.60	0.49	0.56	0.50	0.45	0.50	0.417	0.778	0.208	0.594	0.040**	0.123
Observations	102		99		102		102							

*Notes:* The  $p$ -values reported are from chi-square tests for gender and school and ranksum tests for age and SOEP. Regression analyses control for individual characteristics. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 2.B.2: Number of Groups by Types of Arguments

Number of Groups	Dishonest Arguments			
	Yes	No	Total	
Honest Arguments	Yes	37	5	42
	No	47	37	84
Total	84	42	126	

*Notes:* This table presents the number of groups in which honest and/or dishonest argument is mentioned at least once. Of 135 groups, one group did not chat, and eight groups (two per treatment) viewed the video where ‘☹’ was shown, leaving 126 groups. For ex. in 37 out of 126 groups, there were at least one honest and one dishonest arguments mentioned.

Table 2.B.3: Shares of Dishonest and Honest Arguments

	(1)	(2)	(3)	(4)	(1-2)	(1-3)	(2-4)	(3-4)
	BaseNoPC	BasePC	OathNoPC	OathPC				
	$p$ -value							
Dishonest	61.33%	54.81%	56.39%	43.50%	0.699	0.564	0.403	0.261
Honest	13.67%	9.70%	18.60%	22.12%	0.300	0.407	0.069*	0.819
$N$	32	31	31	32				

*Notes:* This table presents the average shares of dishonest and honest arguments circulated within each group by treatment. Instances where ‘☹’ was shown are excluded. \*  $p < 0.10$

## 2.C Appendix Figures

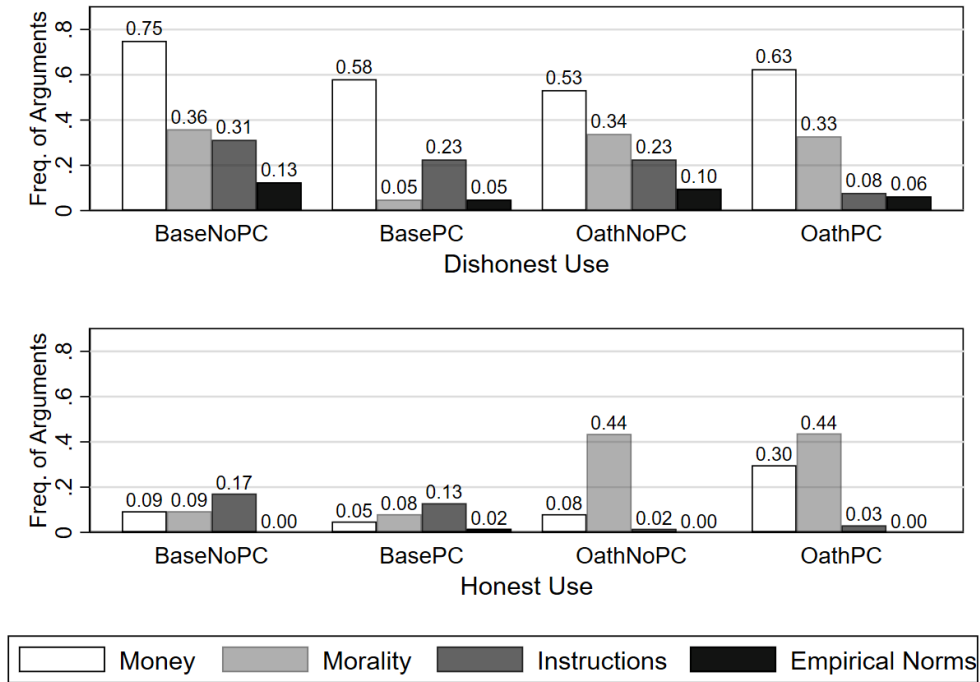


Figure 2.C.1: Frequencies of Each Type of Arguments

*Notes:* This figure displays the frequencies of each type of arguments used in the group chat during part 2 of the experiment. Dishonest use (Upper Panel) Honest Use (Lower Panel).

## 2.D Code Book for Content Analysis (Translated from French)

### Instructions

**Your job is to code messages exchanged by participants in an experiment.**

**Context of the task:** You will read excerpts of conversations exchanged during an experiment that has already taken place. They are messages from a chat where participants communicate with other members of their group. More precisely, there were three members in each group (A, B and C).

Before they could interact with each other, they each saw a video on their screen displaying a roll of the die (1, 2, 3, 4, 5 or 6) that was randomly chosen by the program. Each video had the same probability of being chosen. All three members saw the same video. Their task was to memorize the number displayed on the die so that they could enter it later on their computer. The money they earned depended on the number they entered and was determined as follows:

Number Entered	'1'	'2'	'3'	'4'	'5'	'6'
Tokens Gained	1	2	3	4	5	0

One token was exchangeable for two Euros at the end of the experiment. Before entering their numbers individually, group members were given the opportunity to chat for 5 minutes. Although some of the messages are neutral, they can help you understand the discussion between the group members.

**Data:** For each conversation excerpt, you will have the following information (variable name in parentheses):

1. Group identifier (Group\_id)
2. Subject identifier (Subject\_id)
3. Name in chat (NameInChat → A, B or C)
4. Message sent by the participant to two other members of his/her group (Messages)
5. Time (in seconds), indicating the second the message was sent (TimeStampInChat)
6. Number of the die viewed in the video in part 2 (RandNumOfGroup)

**Coding:** Your job is to sort the messages into different categories based on the arguments participants use to persuade others in the group about the number to enter. You will see examples associated with each category.

**IMPORTANT:** Only count arguments that are EXPLICITLY mentioned. For example: if “3” was shown and the participant said. “I think we should be honest and enter 3” honesty=1 ; “I think we should enter 3” honesty=0.



**List of categories:**

1. **Money\_dishon:** Discussion about money in favor of lying (# of times money/points/chips are mentioned). Ex: *“We should enter 5 because that will maximize our winnings.”*
2. **Money\_hon:** Discussion about money in favor of honesty (# of times money/points/chips are mentioned). Ex: *“We should enter 4, that will still give us 8 euros.”*
3. **Honesty\_dishon:** Discussion about honesty in favor of lying (# of times mentioned). Ex: *“There is no need to be honest; it’s okay to lie.”*
4. **Honesty\_hon:** Discussion about honesty in favor of honesty (# of times mentioned). Ex: *“Let’s be honest and report the number we saw.”*
5. **Insecurity\_dishon:** Discussion of uncertainty in favor of lying (# of times uncertainty or doubt is mentioned). Ex: *“I don’t understand, I thought we could enter any number.”*
6. **Insecurity\_hon:** Discussion of uncertainty in favor of honesty (# of times uncertainty or doubt is mentioned). Ex: *“I don’t understand, I thought we had to enter the number we saw earlier.”*
7. **Rules\_dishon:** Discussion of rules in favor of lying (# of times rules/instructions are referred to). Ex: *“We don’t have to follow the rules; it is not mentioned in the instructions that we can’t lie.”*
8. **Rules\_hon:** Discussion of rules in favor of honesty (# of times rules/instructions are referred to). Ex: *“We have to follow the rules; the instructions say to enter the number seen in the video.”*
9. **Others\_dishon:** Referring to the dishonesty of others outside the group (e.g., the experimenter, people in general, other participants, etc.) (# of times mentioned). Ex: *“Everyone lies; I don’t believe they (the experimenters) showed us a random video.”*
10. **Others\_hon:** Refer to the honesty of others outside the group (e.g., the experimenter, people in general, other participants, etc.) (# of times mentioned) Ex: *“I think the other participants would be honest.”*
11. **Neg\_consqq:** Discussion of negative consequences (# of times they talk about hurting conscience, breaking the oath, fear of punishment). Ex: *“It hurts your conscience if you lie; will we be punished if we break the oath/pledge?”*
12. **Pos\_consqq:** Discussion of positive consequences (# of times they talk about how lying could have a positive outcome). Ex: *“The group gains more if we all lie.”*
13. **Oath\_dishon:** Discussion of the oath in favor of lying (# of times mentioned). Ex: *“The oath is useless.”*
14. **Oath\_hon:** Discussion of the oath in favor of honesty (# of times mentioned). Ex: *“Respect the oath.”*

A sentence can belong to several categories. Ex: *“Let’s be honest and put 3. We will earn 6 euros anyway”* Money\_hon =1 AND Honesty\_hon =1

For each group, you must also note the ‘Subject\_id’ of the first subject who proposed the number to be entered.

## 2.E Additional Analyses

Analyses not made at the time of pre-registration are reported here. They complement the findings in the main text. Section 2.E.1 looks at group coordination on (dis)honest reports. Section 2.E.2 and 2.E.3 examine the beliefs about the empirical norms and decision times under oath, respectively.

### 2.E.1 Group Coordination

This section explores the coordination rates in part 2 of the experiment. [Kocher et al. \(2018\)](#) analysed group coordination under the two incentive schemes. In addition, it is interesting to examine given that [Jacquemet et al. \(2018\)](#) found in the setting of a coordination game that a truth-telling oath increased coordination rates by nearly 50%, resulting in an increase in efficiency because senders' message became more truthful while followers were more willing to trust.

Table 2.E.1 displays the coordination rates at the group level by treatment. The rates in BaseNoPC and BasePC are high (84.4% and 100%, respectively,  $p = 0.053$ , Fisher's exact test).<sup>28</sup> Introducing oath reduces coordination rates, but only when there is no payoff commonality (84.4% in BaseNoPC vs. 56.3% in OathNoPC,  $p = 0.027$ , Fisher's exact test).

Table 2.E.1: Coordination Among Group Members by Treatment

Coordination	Nb. Dishonest	BaseNoPC	BasePC	OathNoPC	OathPC
Yes (all honest)	0	1	2	2	5
No	1	3	0	9	0
No	2	1	0	5	0
No (all dishonest)	3	1	0	0	0
Yes (all dishonest)	3	26	29	16	27
Total Nb. of Groups		32	31	32	32
Coordination rates		84.4%	100%	56.3%	100%

*Notes:* This table presents coordination rates of group members across treatments. Figures are collapsed at the group level. Instances where '☒' was shown are excluded.

Looking at whether groups coordinated on dishonest or honest reports (or remained uncoordinated) yield interesting findings. When group members do not face payoff commonality, oath reduces coordination rates on dishonest reports (81.3% in BaseNoPC vs. 50.0% in OathNoPC,  $p = 0.017$ ), but does not increase coordination rates on honest reports (3.1% in BaseNoPC vs. 6.2% in OathNoPC,  $p = 1.000$ , Fisher's exact tests). This means that when there is no payoff commonality, introducing oath moves people away from dishonest coordination towards respecting their individual honesty oath. This is likely since the rates at which groups remained uncoordinated (*i.e.*, 1 or 2 group members being honest) increase from 12.5% in BaseNoPC to 43.8% in OathNoPC ( $p = 0.011$ , Fisher's exact test).<sup>29</sup> On the other hand, under payoff commonality, oath does not significantly reduce coordination rates on dishonest reports (93.5% in BasePC vs.

<sup>28</sup>These rates are very close to those reported in [Kocher et al. \(2018\)](#): 84.6% in GroupNoPC and 100% in GroupPC.

<sup>29</sup>Among people who remained uncoordinated with their group, 4 of 5 in BaseNoPC were honest

84.4% in OathPC,  $p = 0.426$ , Fisher's exact test) nor does it increase honest coordinations (6.5% in BasePC vs. 15.6% in OathPC).<sup>30</sup>

Thus, an honesty oath breaks down coordination on dishonest reporting when groups do not face payoff commonality – a setting where only pure conformity plays a role and pressure is minimal. On the other hand, the oath has no impact on coordination rates under payoff commonality – a setting where there is group pressure.

These analyses lead to the following observation:

**Observation 1.** Coordination rates in the baseline conditions are close to those reported in [Kocher et al. \(2018\)](#). An honest oath reduces coordination when groups do not face payoff commonality by moving from coordinating on dishonest reporting to remaining uncoordinated. However, when there is payoff commonality, oath has no impact on coordination rates.

### 2.E.2 Beliefs about Empirical Norms

This section explores individuals' beliefs about the empirical norms. Beliefs were elicited at the end of the experiment (before receiving feedback about the payoff). More precisely, for each part of the experiment, participants had to guess the proportion of individuals who reported each die number within the session. The analysis focuses on the participants' belief about the proportion of individuals reporting '5' as it indicates their belief about the level of dishonesty.

#### Belief about Group (Dis)honesty

Overall, participants' belief about the level of group dishonesty is consistent with the impact of oath observed in the behavioral data. Participants in BaseNoPC and BasePC anticipated the proportion of people who reported '5' to be around 79.6% and 85.0% respectively ( $p = 0.113$ , ranksum test). However, when participants were under oath, the beliefs about the level of group dishonesty were 64.2% in OathNoPC and 73.1% in OathPC, which are significantly lower compared to their respective baselines.<sup>31</sup> In addition, the belief about group dishonesty under oath with payoff commonality is significantly higher than that without payoff commonality ( $p = 0.019$ , ranksum test), which is actually observed in the data.

Complementing the above analysis, Figure 2.E.1 displays the average guessed proportion of participants who reported '5' in part 2, conditioned on the fact that they lied or not in part 2. The choice of conditioning is because participants could infer about group dishonesty after the chat (which allowed for inference about what others might do) and what they decided for themselves. The pattern of belief of both honest and dishonest participants is consistent with the behavioral data: individuals under oath anticipated a lower level of group dishonesty compared to those in the baseline, and they believed that the impact of oath is weaker when group members earn a common payoff. However, it is interesting to note that in all treatments, while honest individuals were optimistic about the impact of oath, dishonest individuals were rather pessimistic (Honest vs. Dishonest  $p < 0.0001$  in all treatments, ranksum tests). In a way, dishonest people were more

---

in part 1, while 8 of 19 in OathNoPC. This may imply that these people stick to their personal norms (being honest).

<sup>30</sup>Since none of the groups miscoordinated under payoff commonality, the sum of coordination rates on dishonest and honest reports is equal to 100%.

<sup>31</sup>Using ranksum tests: BaseNoPC vs. OathNoPC,  $p = 0.0001$ ; BasePC vs. OathPC,  $p = 0.001$ .

‘realistic’ as their beliefs were closer to the actual level of dishonesty than honest people. The mean differences between the true proportion and the guess of liars are about 4-7 percentage points, which are significantly smaller than those reported by honest people (between 27 - 39 percentage points, both  $p < 0.0001$  in BaseNoPC and BasePC;  $p = 0.001$  and  $p = 0.003$  in OathNoPC and OathPC, respectively, ranksum tests).<sup>32</sup>

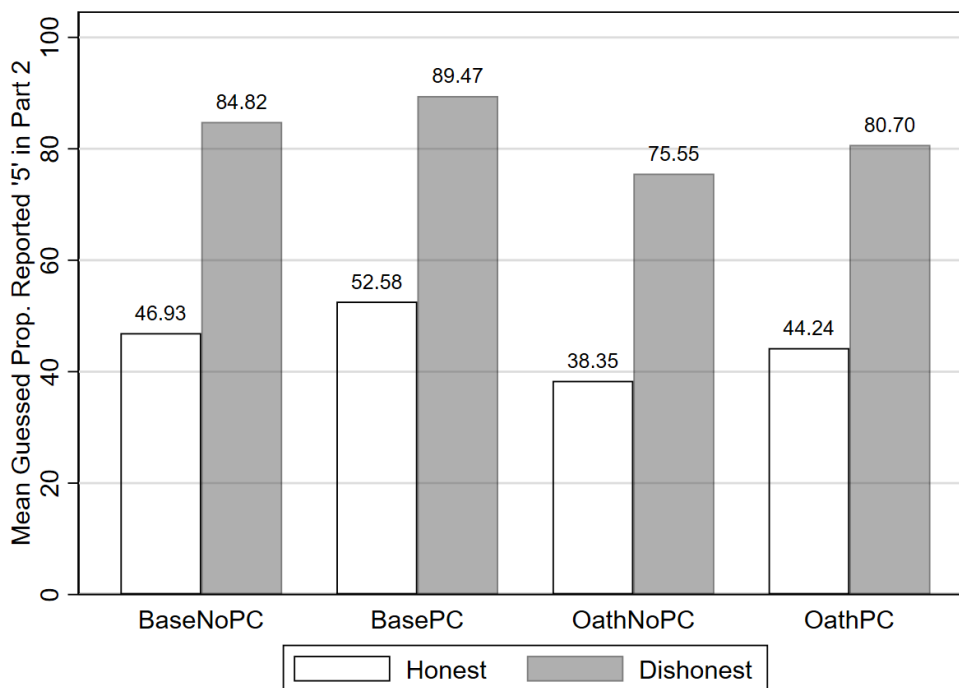


Figure 2.E.1: Belief about Group (Dis)honesty by Type and Treatment

Notes: This figure displays the belief about the proportion of participants reporting ‘5’ in a group setting by treatment, conditioned on their decision in part 2 (honest or dishonest).

### Change in Beliefs about Individual (Dis)honesty

In part 1 and 3 of the experiment, participants played the observed cheating game at an individual setting. Part 3 differs from part 1 in that participants had played the game in a group setting with two other participants in part 2. They had also interacted with them in a group chat via computer. In addition, participants in OathNoPC and OathPC swore to be honest for the rest of the experiment at the beginning of part 2. Given no difference in the die numbers observed in part 1 and 3 in each treatment using two-sided t-tests, if group interactions change the belief participants have about others’ (dis)honesty at the individual level, it can be captured by the difference between the beliefs of others’ (dis)honesty reported for part 1 and 3.<sup>33</sup>

Overall, participants in the BaseNoPC and BasePC became more pessimistic about the level of individual dishonesty. The average change in their belief about the proportion of participants

<sup>32</sup>Excluding those who viewed ‘5’ in part 2 yield the same results.

<sup>33</sup>Change in beliefs is calculated by subtracting the guess about the proportion of participants reporting ‘5’ for part 1 from the guess for part 3. Thus, a positive value means that participants expect the others to become more dishonest after group interaction, and the opposite for a negative value.

reporting ‘5’ is 12.2 in BaseNoPC and 14.5 in BasePC ( $p = 0.661$ , ranksum test). That is, after group interactions, participants anticipated an upward shift in individual dishonesty by about 12-14 percentage points. The shifts in belief are far smaller in the two oath treatments. On average, the change in their belief about the level of individual dishonesty is 3.0 in OathNoPC and 6.1 in OathPC, which are significantly smaller than their respective baselines.<sup>34</sup> This means that while participants in the baseline conditions anticipated group interactions to make people become more dishonest in part 3, those in OathNoPC and OathPC believed that people would be, more or less, as (dis)honest as before the group interactions and oath-taking. This is likely because the effects of group interactions and oath on dishonesty are opposing each other.

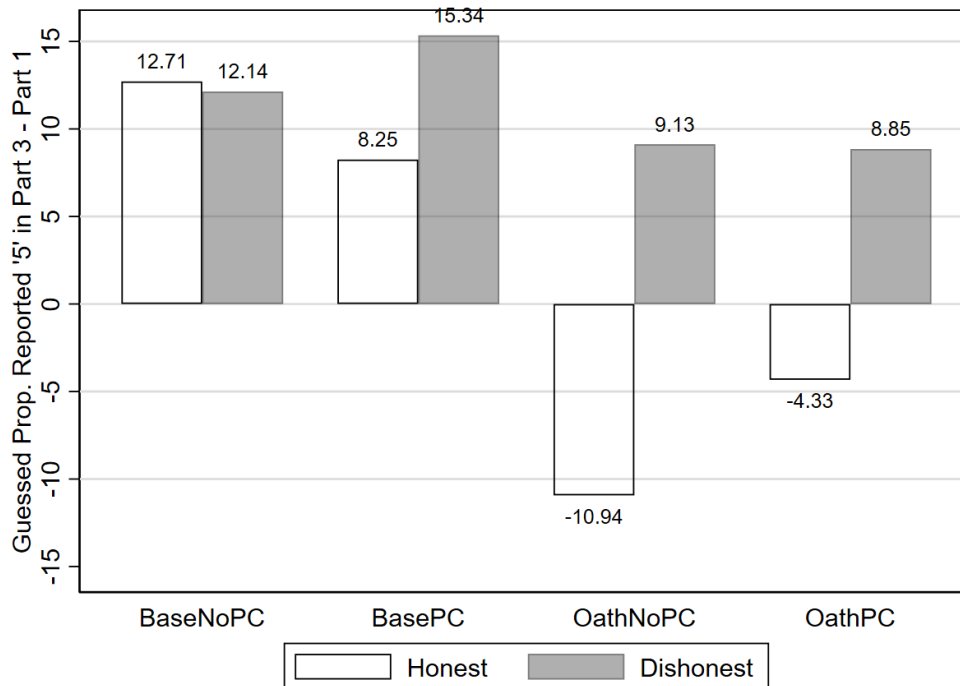


Figure 2.E.2: Change in Beliefs about Individual Honesty After Group Interactions

*Notes:* This figure displays the change in beliefs about the proportion of participants reporting to ‘5’ in an individual setting after group interactions in part 2 by treatment, conditioned on the decision in part 2 (honest or dishonest).

To complement the above analysis, Figure 2.E.2 displays the average change in belief about the proportion of individuals reporting ‘5’ in part 1 and 3, conditioned on the fact that they lied or not in part 2.<sup>35</sup> This explains why the changes in belief are smaller in OathNoPC and OathPC. Those who were honest and dishonest in the group setting in BaseNoPC and BasePC tended to update their belief about individual dishonesty upward (that is, expecting higher level of dishonesty after group interaction). The patterns look different in OathNoPC and OathPC—honest people revised their belief about individual dishonesty downward (that is, expecting higher level of honesty after group interactions) due to the introduction of oath. However, liars in the

<sup>34</sup>Using ranksum tests: BaseNoPC vs. OathNoPC,  $p = 0.0002$ ; BasePC vs. OathPC,  $p = 0.002$ . There is no difference between OathNoPC and OathPC,  $p = 0.636$ .

<sup>35</sup>As done for the analysis of the belief about group dishonesty, this choice of conditioning is because participants might adjust their beliefs of individual dishonesty after having interacted with others in their group and what they decided themselves.

group setting in OathNoPC and OathPC updated their belief about individual dishonesty upward in a similar fashion to those in the baseline conditions.<sup>36</sup>

Finally, comparing the pooled baseline conditions to pooled oath treatments, the difference in the change in beliefs about individual dishonesty is significant both for honest and dishonest people ( $p < 0.0001$  and  $p = 0.009$ , respectively, ranksum tests).<sup>37</sup> This means that without any oath, people anticipate group interactions to drive individual dishonesty upwards. On the other hand, oath leads to a lower anticipation of individual dishonesty, which comes from honest people being optimistic, while dishonest people being *less* pessimistic about others' dishonesty.

These analyses lead to the following observation:

**Observation 2.** Overall, beliefs about group dishonesty is consistent with actual behavior. People anticipate oath to reduce group dishonesty, and that the impact of oath is larger when group members do not face the payoff commonality rule. Without oath, people anticipate group interactions to shift individual dishonesty upwards. When oath is introduced, however, people anticipate a lower shift in individual dishonesty after group interactions, resulting from honest people being more optimistic and dishonest people being less pessimistic.

### 2.E.3 Decision Times Under Oath

This section explores the time participants spent before reaching a decision in part 2.<sup>38</sup> There are two opposing psychological theories regarding the decision time and lying behavior (see Köbis et al. (2019) for a meta-analysis). The first posits *intuitive honesty* on the grounds that lying is psychologically costly. Given that one already knows the truth, truth-telling should be quick and intuitive, while lying is deliberate and requires time. In contrast, the other theory asserts *intuitive dishonesty*, by which truth-telling requires self-control and resistance to temptations. This means that lying is an intuitive response, while truth-telling is a slow reaction.

In part 2 of the experiment, participants interacted with group members before making the decision. Thus, they were exposed to the arguments in favor of honesty and/or dishonesty. Furthermore, as an oath should induce a higher moral cost of lying, this would mean that individuals under oath might need time to deliberate through the moral dilemma.

Table 2.E.2 displays the average response time (in seconds) in part 2 by treatment. Instances where '☒' was shown are excluded so the value reflects the time spent prior to deciding whether to lie or not. Overall, participants spent on average 8-10 seconds in the baseline conditions ( $p = 0.345$ , ranksum test). Introducing oath increases the response time, but only when group members did not earn a common payoff. Participants in OathNoPC spent about 14 seconds while those in the OathPC spent about 8 seconds ( $p < 0.001$ , ranksum test). This could mean that participants in OathNoPC needed more time due to the conflict between morality and the arguments made by their group members. In contrast, in OathPC, as group members must have already reached an

<sup>36</sup>Using ranksum tests: Honest vs. Dishonest; OathNoPC,  $p < 0.0001$ ; OathPC,  $p = 0.004$ . Excluding those who saw '☒' in part 2 yield the same results.

<sup>37</sup>Excluding those who saw '☒' in part 2 yield the same results.

<sup>38</sup>After the group chat, participants had 180 seconds to enter the die number. When hitting the button 'Submit', z-Tree records the time remaining displayed on the decision screen. Subtracting this value from 180 seconds can be used as a proxy of response time the participant spent in the decision-making. In the event that participants exceeded the allotted time, z-Tree records the value as '99999'. This is replaced by '-1', such that the response time is proxied at 181 seconds.

agreement in the chat, the response time remains unaffected as a result of their accord and possibly due to the shared justification to lie.

Table 2.E.2: Response Time (in seconds) in Part 2

	(1)	(2)	(3)	(4)	(1-2)	(1-3)	(2-4)	(3-4)
	BaseNoPC	BasePC	OathNoPC	OathPC	<i>p</i> -value			
All	10.49 (12.29)	8.18 (4.45)	14.41 (18.32)	8.02 (5.97)	0.345	0.087*	0.300	0.0003***
<i>N</i>	96	93	96	96				
Low States	11.26 (15.04)	8.67 (4.88)	16.77 (22.21)	8.15 (6.22)	0.752	0.042**	0.263	0.0005***
<i>N</i>	57	57	57	60				
High States	9.36 (6.48)	7.42 (3.58)	10.95 (9.55)	7.81 (5.61)	0.246	0.801	0.771	0.161
<i>N</i>	39	36	39	36				
Liars	10.06 (10.67)	8.25 (4.48)	12.89 (15.06)	8.06 (5.99)	0.320	0.301	0.322	0.004***
<i>N</i>	88	87	71	81				
Non-liars	15.25 (24.62)	7.17 (4.12)	18.72 (25.31)	7.80 (6.05)	0.944	0.151	0.984	0.027**
<i>N</i>	8	6	25	15				

*Notes:* This table presents mean response time (in seconds) in part 2 of the experiment. Standard deviations in parentheses. Instances where ‘☒’ was shown are excluded. Figures are reported for all participants and separately for the following subsamples: those who observed low states (0, 1 or 2 points), those who observed high states (3 or 4 points), those who decided to lie and those who did not lie. *p*-values reported using ranksum tests. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Looking at the response time conditional on the die numbers observed reveals a clear effect of oath when group members do not earn a common payoff. When states are low (*i.e.*, 0, 1 or 2 points), participants in OathNoPC spent about 17 seconds before reaching the decision and this is significantly longer than those in BaseNoPC (11.26 seconds,  $p = 0.042$ ) and OathPC (8.15 seconds,  $p < 0.001$ ). This means that when there is an economic reason to cheat, oath increases the decision times but only when group members do not earn a common payoff. Under payoff commonality, the decision time conditioned on low states is not affected by the oath ( $p = 0.263$ , ranksum tests). The response times under high states are similar across treatments. Finally, restricting to only those who eventually lied, those in OathNoPC spent about 13 seconds and this is significant longer than those in OathPC (8 seconds,  $p = 0.004$ ). Non-liars in OathNoPC spent about 19 seconds to decide and this is significantly longer than their counterparts in OathPC who spent about 8 seconds ( $p = 0.027$ , ranksum tests).

In sum, the analysis of decision times is consistent with the main finding that the impact of oath against lying extends to a group setting where there is no payoff commonality. It suggests that oath increases the time required to reach a decision whether or not to lie, but only when group members do not face the payoff commonality rule. The increase is pronounced when there is an economic motive to lie. In addition, the increased response time is reflected in both dishonest and honest participants. Longer response time in OathNoPC could result from a deliberation process imposed by the heightened conflict between morality and group communication. Nonetheless, oath does not affect the decision time when group members earn a common payoff, possibly because the rule requires group members to coordinate on their reporting. Besides, having already agreed on the number to report could create a shared sense of justification in misreporting the number. Finally, in comparison to the psychological theories of decision time (Köbis et al., 2019), the data in this experiment seem to be more in line with dishonesty being intuitive: in BaseNoPC and

OathNoPC, truth-tellers seem to take longer time to decide than liars.

These analyses lead to the following observation:

**Observation 3.** Individuals under oath spend significantly more time in making their decision whether or not to lie, but only when there is no payoff commonality rule. Oath does not affect the decision time when group members earn a common payoff.





# Chapter 3

## Risk-Taking with Unethically Earned Money<sup>1</sup>

### 3.1 Introduction

The reputation of the financial sector in the recent decade has been tarnished by the prevalence of banking scandals and frauds revealing a lack of ethical standards and excessive risk taking with client's money (Guiso et al., 2008; Sapienza and Zingales, 2012; Kantšukov and Medvedskaja, 2013). According to a study by Egan et al. (2018), approximately 7% of financial advisors in the US between 2005 and 2015 have misconduct records, and this goes up to 15% in some large financial firms. While some studies showed that individuals tend to take more risk when making decisions on behalf of others (Andersson et al., 2014; Vieider et al., 2016; Füllbrunn and Luhan, 2017), there might be another phenomenon - people may take more risk with money that has been earned unethically. This could be the case if such money was treated as an unexpected windfall money, which may matter since it has been shown that advisors' own preferences influence clients' portfolio allocation (see Foerster et al., 2017; Kling et al., 2019) and they invest in personal portfolio just like they advise their clients (Linnainmaa et al., 2021). However, the recent behavioral economic literature on dishonesty has emphasized the importance of moral costs in ethical behavior (*e.g.*, Fischbacher and Föllmi-Heusi, 2013; Gneezy et al., 2018; Abeler et al., 2019). If such moral costs are important, then unethical earnings may be treated rather like hard-earned money. Therefore, the implications of earning money unethically *vs.* ethically on future risk taking are *a priori* unclear.

In this chapter, we investigated if individuals make different risky decisions when the bad outcome of such decisions reduces the money previously earned unethically

---

<sup>1</sup>This Chapter is joint with Marie Claire Villeval.

or ethically through effort or luck. The principle of economic fungibility posits that any unit of money is substitutable. This implies that individuals' consumption decisions should only be influenced by the total wealth, and not by its composition. This, however, has been refuted by studies in mental accounting (Thaler, 1985, 1999; Abeler and Marklein, 2017) which showed that the source of earnings influences how they are spent. This encompasses the notion of labels associated with the money (*e.g.*, a bonus, a grant, or a subsidy) and how it has been earned (through effort or luck). We extend this reflection by studying whether the unethical *vs.* ethical source of earnings can also lead to a violation of the principle of fungibility of money.

Imas *et al.* (2020) have explored the violation of the fungibility of unethical money in the context of charitable giving. We differ from them in that we investigated such a violation in the context of risk taking. Moreover, they studied a motivated violation through a mechanism of mental money laundering that is engaged by dishonest individuals to dissociate unethical money from its source by exchanging it with 'clean' money from another source. In contrast, by studying how much risk individuals are willing to take with dishonestly *vs.* honestly earned money, we explored a non-motivated violation based on mechanisms identified when studying violations of fungibility between money earned through luck or effort.

The observation that individuals tend to use money earned through costly effort differently from a windfall gain (*i.e.*, Hoffman *et al.*, 1994; Cherry *et al.*, 2002; List and Cherry, 2008; Hvide *et al.*, 2019) has been explained by a sense of entitlement because effort provision imparts a feeling of ownership. In the context of risk taking, a sense of entitlement should induce risk aversion. In contrast, obtaining a windfall gain through luck imparts a weaker sense of entitlement, which should induce higher risk taking. However, how would a person take risk with unethical money? The answer to that question is not straightforward for two reasons.

On the one hand, when money is earned dishonestly by deceiving or lying to someone, in principle it should not lead to any feeling of entitlement because this money objectively belongs to others. One may think of it as something easily earned, like a 'house money' (Thaler and Johnson, 1990).<sup>2</sup> This may imply that a dishonest individual perceives unethical money as a windfall gain that, thereby, induces risk taking. On the other hand, the individual may incur a moral cost to obtain unethical money, due to intrinsic lying aversion or reputational concerns associated with perceived cheating aversion (Dufwenberg and Dufwenberg, 2018; Gneezy *et al.*, 2018; Abeler *et al.*, 2019; Khalmetski and Sliwka, 2019). Such moral costs might

---

<sup>2</sup>We are talking here of dishonest acts that are typically associated with the exploitation of asymmetric information. We are not considering criminal activities that entail other types of costs than psychological costs.

act like an effort cost which, then, induces risk aversion.<sup>3</sup> In other words, dishonest individuals may justify a sense of entitlement to the unethical gain and thus, behave *as if* they were actually entitled to it. This reasoning is based on the psychological cost hypothesis of [Thielmann and Hilbig \(2019\)](#) who argued that although dishonest individuals are not *objectively* entitled to the unethical gain, they may justify *subjectively* that it is ‘*theirs*’ because they have incurred a moral cost to obtain it. If this is so, suffering a moral cost to obtain this money may reduce the willingness to put it at risk in future decisions.

Whether individuals treat dishonest money as a windfall gain or as an earned income when making risky decisions with this money would be difficult to identify by using natural data from the field. First, dishonest behavior is usually hidden. Second, even it was possible to measure the risky behavior of honest *vs.* dishonest individuals without noise, the treatment of honest *vs.* dishonest monetary currencies would be impossible to disentangle. Therefore, we designed an online experiment in which participants started by earning money and, then, made a risk-taking decision whose good outcome would increase the preliminary earnings, whereas bad outcome would reduce these earnings. This setting also allows us to have a perfect control on earnings and risk opportunities.

Our between-subjects design consists of three two-stage treatments that varied in how the participants earned money in the first stage. In the *Lying treatment*, participants played the truth-telling mini-game of [Gibson et al. \(2013\)](#) in which they could lie by misreporting an information to earn more money. In the *Effort treatment*, participants performed a real-effort task to earn money. In the *Windfall treatment*, participants earned money through luck in a binary lottery. In the three treatments, earnings could be either low or high, but their amounts were held similar across treatments. In the second stage of each treatment, participants performed a variant in the loss domain of the Bomb Risk Elicitation Task (BRET) introduced by [Crosetto and Filippin \(2013\)](#). In the event of a good outcome (*i.e.*, participants did not select the box hiding a bomb), their earnings would increase in the number of boxes collected, whereas in the event of a bad outcome (*i.e.*, they selected the box hiding a bomb), they would lose a fraction of the income earned in the first stage of the experiment.

We conjectured that, for given baseline risk preferences, if the moral cost of lying generates a sense of entitlement, as posited by [Thielmann and Hilbig \(2019\)](#), there should be no or little difference in risk taking in the BRET between individuals whose high earnings were realized through lying and those whose same earnings resulted

---

<sup>3</sup>This effect might be even stronger when dishonest acts require a technology to be successful and reduce one’s risk of detection and sanction.

from effort, because both groups would exhibit a sense of entitlement. Along this line, we also conjectured that the former should take less risk than the individuals whose same earnings were realized through luck, because the former would exhibit a sense of entitlement but not the latter.

Contrary to these conjectures, our results showed that, conditional on their baseline risk preferences (as self-reported in the SOEP questionnaire), dishonest individuals took more risk in the BRET than those who earned the same earnings through effort. Additionally, they revealed no significant difference in risk taking between individuals who earned money by lying or luck, while individuals took more risk with money earned by luck than through effort provision. These results imply that dishonest individuals treated unethical money more like a windfall gain than hard-earned money. This could either mean that the moral cost of lying does not generate a feeling of entitlement, in contrast with costly effort, or that liars were individuals who self-selected based on their low moral cost of lying. The second interpretation is plausible because the choice to lie in our experiment was observable *ex post* to the experimenter.

To disentangle between these two interpretations, we conducted a follow-up experiment in which we increased the moral cost of lying *ex-post*. We did so by informing the participants about the injunctive norm regarding the appropriate behavior in the truth-telling mini-game of [Gibson et al. \(2013\)](#), as elicited in a separate study, after participants made their lying decision but before they performed the BRET. The results of the *Higher Moral Cost treatment* showed that dishonest individuals took significantly less risk in the BRET than those who participated in the original Lying treatment. This finding is in line with the psychological cost hypothesis that liars may justify their entitlement to the unethical gain, based on the moral cost they incurred. The feeling of entitlement to unethical gains may thus be sensitive to the level of the moral cost.

Exploring the heterogeneity of our effects, we found that the increased willingness to take risk with unethical money in the original Lying treatment was driven by individuals with risk averse baseline preferences. Risk averse liars were taking higher risk than those with similar baseline risk attitudes who were taking risk with other sources of earnings. This tendency disappeared when the moral cost of lying was increased in the new treatment. We interpret such higher risk taking with unethical gain as a house money effect: since this gain was obtained easily, individuals were willing to put it at stake as if it had no effort value. In contrast, when the moral cost of lying increased, a feeling of entitlement may have developed.

Overall, our findings highlight the importance of taking into account both the unethical source of earnings and the moral costs associated with unethical behavior

on risk taking decisions. They suggest that increasing the moral cost of dishonesty, notably through reminders of the injunctive moral norms, leads individuals to have a more conservative use of money in the domain of risk. These findings may be relevant notably in the financial industry where fraud is not uncommon and where the decision-making environment already tends to encourage excessive risk taking. In addition, [Egan et al. \(2018\)](#) reported that about 44% of advisors who were fired for misconducts found a new job in the same year. Moreover, there was a matching market for misconducts – advisors with misconduct records sought out firms that hired people like them, while these firms were less likely to fire advisors for new misconducts. Our results suggest that if the moral cost of obtaining unethical money is low, it then implies to a tendency to take risk, which can indirectly influence the way advisors allocate portfolio for their clients (see [Foerster et al., 2017](#); [Kling et al., 2019](#); [Linnainmaa et al., 2021](#)).

The remainder of this chapter is organized as follows. Section 3.2 briefly reviews the related literature. Sections 3.3 and 3.4 outline the design and the behavioral conjectures. Section 3.5 reports the results of our main experiment. Section 3.6 describes our follow-up experiment with increased moral cost of lying and its findings. Section 3.7 discusses these findings and concludes.

## 3.2 Related Literature

This chapter is connected to three strands of literature studying, namely, the fungibility of money, the contingency of risk attitudes, and the spillover effects of dishonest behavior on future decisions.

First, the violation of fungibility of money has been corroborated by evidence from both field (*e.g.*, [Beatty et al., 2014](#); [Kooreman, 2000](#); [Hastings and Shapiro, 2013, 2018](#)) and laboratory experiments. In particular, [Abeler and Marklein \(2017\)](#) observed the spending behavior of restaurant customers by giving them a coupon, either earmarked for beverages or unlabeled (*i.e.*, that could be spent on both food and beverages). They found that customers receiving a coupon earmarked for beverages spent more on beverages than their counterparts who received an unlabeled coupon. Other violations of fungibility come from different feelings of ownership, as shown by studies on the endowment effect in economic games. [Hoffman et al. \(1994\)](#) found that letting individuals earn the right to be the proposer and dictator in the ultimatum and dictator games, respectively, led to a significant reduction in the generosity of transfers because individuals perceived to have earned the rightful entitlement of the endowment (see also [Cherry, 2001](#); [Cherry et al., 2002](#); [List and](#)

Cherry, 2008; Oxoby and Spraggon, 2008). Having to earn one's endowment creates a sense of entitlement compared to when it is a windfall money, which impacts future decision making with this endowment.

We contribute to this reflection on violations of fungibility by studying whether risk taking differs depending on whether the money at stake has been earned by chance, through effort, or dishonestly through lying. Few studies investigated how individuals perceive money earned from an unethical source and how its nature affects its usage. A psychology study by Tasimi and Gelman (2017) found that compared to money earned honestly, individuals find morally tainted money less desirable, are less likely to spend it on themselves, and prefer to donate it to a charity.

The closest study to ours is that of Imas et al. (2020) which explored the tendency of dishonest individuals to engage in motivated mental money laundering. The money earned in a Sender-Receiver game was wired through a lottery that would return the same amount with a probability of 83%, and double or lose with a probability of 8.3% each. In the laundered condition, the money wired through the lottery was replaced with physically different bills from a clean source, whereas in the unlaundered condition, there was no such physical exchange. After the lottery stage, senders participated in a donation game. Imas et al. (2020) found that dishonest senders whose bills were exchanged donated significantly less than those whose bills were not exchanged. In an additional experiment, they found that, when allowed to choose the amount to be wired, dishonest senders whose bills would be exchanged put significantly more money at stake than those whose bills would not be exchanged. They argued that dishonest senders exploited the opportunity to launder money, even at some risk of losing it, which violates the principle of fungibility.

Although individuals took risk with dishonestly earned money in both studies, we differ from Imas et al. (2020) in two major respects. First, we study risk taking, which is not related to the moral domain, whereas Imas et al. (2020) studied moral motivation through money laundering and charitable giving. Second, our risk task is very different from that used in Imas et al. (2020). Their lottery was designed to launder money (*i.e.*, it offered a high chance of receiving the same amount back in 'clean' bills). In contrast, our participants decided how much risk to take, knowing that the negative outcome of the lottery would reduce the money previously earned, while the positive outcome would add gains to the participants' dishonestly earned money but not substitute for it. Our contribution to the literature on fungibility is investigating how the unethical nature of money may influence how individuals use it in the context of risk.

Second, this chapter relates to the literature on the contingency of risk attitudes. Previous studies showed that the willingness to take risk depends on prior outcomes. Some found evidence that risk taking increases after gains (*e.g.*, [Ackert et al., 2006](#); [Corgnet et al., 2015](#); [Suhonen and Saastamoinen, 2017](#)), which is consistent with the house money effect ([Thaler and Johnson, 1990](#)). Others found evidence that risk taking decreases after gains but increases after losses, which supports the disposition effect ([Shefrin and Statman, 1985](#); [Weber and Camerer, 1998](#)). [Imas \(2016\)](#) showed that a prior loss reduces risk taking when it is realized, but increases risk taking when it is a paper loss. Other studies examined the effect of stakes (*e.g.*, [Holt and Laury, 2002, 2005](#); [Deck et al., 2008](#); [Lefebvre et al., 2010](#); [Fehr-Duda et al., 2010](#)), emotions (*e.g.*, [Nguyen and Noussair, 2014](#); [Cohn et al., 2014](#); [Guiso et al., 2018](#)), and risk taking for others (*e.g.*, [Andersson et al., 2014](#); [Vieider et al., 2016](#); [Füllbrunn and Luhan, 2017](#)).

We add to this literature by investigating whether the unethical nature of previous gains that could be put at stake matters for the propensity to take risk. In general, individuals decide more conservatively with earned money than with ‘unearned’ money ([Thaler and Johnson, 1990](#); [Arkes et al., 1994](#)). [Hvide et al. \(2019\)](#) found that individuals who had to earn their endowment from a real-effort task took less risk than those who simply received an endowment. Closer to our research question, a psychology study by [Chen et al. \(2017\)](#) explored how the willingness to take risk is influenced by the ethical nature of money. They found that forced liars in a deceptive task were more likely to choose a risky option in a latter task. They argued that it is a feeling of guilt that caused them to take more risk as a distancing strategy. The cognitive dissonance arising from a misalignment of attitudes and actions of individuals with an honest principle who were forced to lie may have induced a disentanglement effect of the unethical gain. This is opposite to the psychological cost hypothesis of [Thielmann and Hilbig \(2019\)](#) mentioned earlier. This chapter differs from [Chen et al. \(2017\)](#) in that we looked at the situation where individuals decided to lie (and therefore were responsible for the lie they told) rather than being forced to, and we apply standard procedures in experimental economics to examine the tendency to take risk with that money.

Lastly, this chapter relates to the investigations of the spillover effects of dishonest decisions on future behavior in an unrelated activity. The theory of self-concept maintenance ([Mazar et al., 2008](#)) postulates that individuals have an innate desire to perceive themselves in a positive light. One phenomenon that supports the theory is moral cleansing ([Monin and Miller, 2001](#); [Shalvi et al., 2015](#)), whereby individuals who committed an immoral action are more likely to engage in moral behaviors in the future to restore a positive self-image. [Gneezy et al. \(2014\)](#) found that liars in a



Sender-Receiver game were more likely to donate money to a charity. A related concept is moral balancing whereby individual keeps checks and balances of good and bad actions (*e.g.*, [Mazar and Zhong, 2010](#); [Ploner and Regner, 2013](#); [Brañas-Garza et al., 2013](#); [Rahwan et al., 2018](#)). Although we also contribute to the understanding on how earning money dishonestly influences a future decision, we differ from these studies in that we focus on risk taking, which is unrelated to the moral domain. We are not exploring a behavioral spillover but how behavior influences the status of the money earned and how, in turn, this status influences future risk taking.

### 3.3 Experimental Design and Procedures

#### Design

The experiment consists of two parts following a preliminary questionnaire. In part 1, participants generated an endowment. The design of part 1 differs across three between-subjects treatments (Lying, Effort and Windfall) that varied how the endowment was earned. In part 2, common to all treatments, participants performed a risk task in which a fraction of the endowment earned previously could be put at stake. Instructions (available in Section 3.A of the Appendix) were distributed at the beginning of each part.

#### Elicitation of Baseline Risk Preferences

At the beginning of each session, we collected two self-reported measures of risk attitudes taken from the German Socio-Economic Panel (SOEP) ([Dohmen et al., 2011](#)). Participants answered the following question on a 0-10 scale: “How do you see yourself: are you generally a person who is fully prepared to take risk or do you try to avoid taking risks?”. We collected this measure of risk in general and in the financial domain. We used them to control that any difference observed across treatments was not driven by differences in baseline risk attitudes or by different self-selection rules in terms of risk attitudes.

#### Part 1: Formation of Earnings

Participants generated a \$1 or \$2 endowment that could be put at stake in the risk task in part 2 (unknown from the participants when they performed part 1). How the endowment was earned depended on the treatment assigned.

**Lying Treatment:** Participants played the truth-telling mini-game of [Gibson et al.](#)

(2013). They received the scenario of a CEO who had to announce earnings per share for the previous quarter. They were told that, as a CEO, their compensation depended on the earnings they announce. They were also told that the market currently anticipated the announcement of 35 cents per share, but the true earnings was 31 cents per share. They were informed that they would be paid based on the CEO's compensation (their decision). Participants earned \$1 by choosing to announce 31 cents per share (telling the truth), or \$2 by announcing 35 cents per share (telling a lie).

This task allows us to identify the cheaters at the individual level (a lie is observable *ex-post*) and luck does not play any role. There is evidence in the literature that choosing a dishonest option in the CEO task is perceived as socially inappropriate (see Huber and Huber, 2020). Alternative cheating games with no identifiability (*e.g.*, Fischbacher and Föllmi-Heusi, 2013; Cohn et al., 2014; Abeler et al., 2019) would have the advantage of a lower reputation concern, but they could give rise to interpretation issues. Indeed, participants reporting a high payoff outcome could have got lucky and, thereby, did not earn the money through lying. Any observed effect in our setting could then have been an artifact of getting lucky in disguise of 'lying'.<sup>4</sup>

**Effort Treatment:** In the Lying treatment, participants could earn \$1 or \$2, depending on their decision in the CEO task. Therefore, we similarly implemented two levels of earnings in the other treatments. In the Effort treatment, participants performed an Encoding task (*e.g.*, Erkal et al., 2011; Gangadharan et al., 2017). They received a table with letter-number pairs and were asked to encode as many words as possible within eight minutes by entering the number corresponding to each letter of each given word (see screenshots in the instruction in Section 3.A of the Appendix). They earned \$2 if they correctly encoded at least 39 words; otherwise, they earned \$1. We chose this threshold such as to have about 50% of participants with a high or a low endowment, like in the other treatments.

**Windfall Treatment:** In the Windfall treatment, participants did not have to perform any task. Instead, they were informed that they would receive either \$1 or \$2 with an equal chance.

---

<sup>4</sup>Repeating the task, as in Cohn et al. (2014), to identify cheaters at the individual level was not an option because it could have created several levels of wealth (in case of paying all rounds) or allowed a self-serving interpretation of the payoff (in case of paying one randomly selected round) since even dishonest participants would have actually observed the high payoff outcome in some rounds.

## Part 2: Risk Taking Task

In part 2, participants performed the Bomb Risk Elicitation Task (BRET) developed by [Crosetto and Filippin \(2013\)](#), adapted in the loss domain, whereby a fraction of the endowment generated in part 1 could be put at stake. The adapted BRET represents a set of 100 lotteries with the two extremes being the degenerate lotteries with no uncertainty, while the remaining 98 lotteries constitute mixed gambles.

Precisely, participants were presented a 10x10 grid, each cell representing a box. They had to decide how many boxes they were willing to collect (0-100 inclusive). For each box collected, they could earn \$0.03. However, a bomb was hidden behind one of the boxes. Each box was equally likely to contain the bomb. The position of the bomb was randomly determined by the program and would be revealed at the end of the session. The payoff depended on whether the bomb had been collected or not. If the bomb had not been collected, participants earned the payoff from collecting the boxes, in addition to the gain of part 1. If the bomb had been collected, they earned nothing from collecting the boxes *and* they lost \$0.5 from their endowment.<sup>5</sup> Before making their decision, participants responded to some comprehension questions, then they played one practice round to familiarize with the interface. After the practice round, they chose the number of boxes they were willing to collect.

We departed from the standard BRET in two ways. First, following [Gioia \(2017\)](#), players provided the number of boxes they wished to collect. The program collected the boxes in numerical order from the top left corner. Participants could modify their choice as they wished before validating the decision (see screenshots in the instruction in Section 3.A of the Appendix). Compared to pushing a stop button or manual selection, this version minimizes boredom or impatience that could introduce noise in the decision. Second, while in the standard BRET collecting a bomb nullifies the earnings from the task, in our design it also destroys a fraction of the player's initial earnings. This is similar to the variant used in [Nielsen \(2019\)](#) in which the endowment is lost if a bomb is collected.

## Procedures

1,048 individuals, recruited in the United States through Amazon Mechanical Turk, participated in the experiment.<sup>6</sup> 373 participated in the Lying treatment, 372 in the

---

<sup>5</sup>Thus, the set of available lotteries was (\$0, 1), (\$0.03, 0.99; -\$0.5, 0.01), (\$0.06, 0.98; -\$0.5, 0.02), (\$0.09, 0.97; -\$0.5, 0.03), (\$0.12, 0.96; -\$0.5, 0.04), ... , (\$2.88, 0.04; -\$0.5, 0.96), (\$2.91, 0.03; -\$0.5, 0.97), (\$2.94, 0.02; -\$0.5, 0.98), (\$2.97, 0.01; -\$0.5, 0.99), (-\$0.5, 1).

<sup>6</sup>The number of observations is based on a pre-registered power calculation. Assuming a type-I error rate of 0.05, a power level of 0.8 and a small effect size (Cohen's  $d = 0.3$ ), the required number

Windfall treatment, and 303 in the Effort treatment. We excluded ten individuals who did not pass the attention checks and five individuals who collected 100 boxes in the BRET, revealing a misunderstanding of the task. Therefore, we are left with 1,033 observations (371 in the Lying, 364 in the Windfall and 298 in the Effort treatments). The summary statistics of the participants' individual characteristics are shown in Table 3.B.1 in Appendix. They do not differ across treatments, in particular for the baseline risk attitudes, except for age in the Effort and Windfall treatments. Participants in the Effort treatment ( $\text{Mean}_{\text{Age}} = 42.47$ , Standard Deviation (SD) = 12.87) were slightly older than those in the Windfall treatment ( $\text{Mean}_{\text{Age}} = 40.58$ , SD = 12.88;  $p = 0.043$ , Mann-Whitney U test).

At the end of the session, participants learned about the location of the bomb in the BRET and answered a short questionnaire and attention checks. They received a fixed payment of \$0.50, plus a variable payment depending on their earnings in part 1, and the outcome of the BRET in part 2. They earned on average \$2.38 (SD = 0.82). The average duration of the experiment was 10 minutes (SD = 5.88).

A total of 556 individuals earned \$2 in part 1 of the experiment: 194 in the Lying treatment, 182 in the Windfall treatment, and 180 in the Effort treatment. The proportions of individuals who earned \$2 in the Lying and Windfall treatments (52.3% *vs.* 50.0%, respectively) are not statistically different ( $p = 0.534$ , chi-square test), but they differ between the Lying and Effort treatments (52.3% *vs.* 60.4%, respectively;  $p = 0.036$ ) and between the Windfall and Effort treatments ( $p = 0.007$ ). Their characteristics are summarized in Table 3.B.2 in the Appendix; almost none are significantly different across subsamples.

### 3.4 Conjectures

In this section, we formulate two behavioral conjectures regarding individuals' risk taking in the BRET conditional on the type of money at stake, controlling for their baseline risk attitudes and the amount of their earnings.<sup>7</sup> To answer our research

---

of observations to uncover the hypothesized effect between the Lying and Windfall treatments is 184 observations per treatment. No existing study gives a direct insight on the lying rate in the CEO task with a MTurk sample. In [Gibson et al. \(2013\)](#), the average lying rate (excluding no lying incentive) among student samples was 68%. The pilot conducted yielded a similar lying rate at 63%. Given this, we conservatively assumed a lying rate of 60% and set the required number of observations per treatment at 307, resulting in a total number of observations of 921 (307x3). However, contrary to this initial power calculation, the empirical lying rate turned out to be only 50%. Given this lower lying rate, we collected about 65 additional observations for the Lying and Windfall treatments each, to uncover the hypothesized effect.

<sup>7</sup>The design and behavioral conjectures have been pre-registered at AsPredicted (#49535).

questions, we focus on the behavior of the participants who earned a \$2 endowment in the different treatments because only those who received \$2 in the Lying treatment earned their endowment through lying, while those who received a \$1 endowment earned it honestly.

We first compare risk taking in the BRET in the Effort and Windfall treatments to establish a benchmark of the entitlement effect. As shown by the studies cited in Section 3.2, this effect posits that individuals are more protective of an endowment generated through effort rather than luck, because of a stronger sense of entitlement. In the context of our experiment, we anticipated that participants in the Effort treatment should take less risk than those in the Windfall treatment because the former group earned their endowment through effort provision, while the latter group earned it through luck.

The comparison of risk taking in the Lying and Effort treatments allows us to identify whether a moral cost and a cognitive cost both lead to a more conservative behavior in the BRET. In the Lying treatment, participants earned \$2 dishonestly in the CEO task, at a possible moral cost. Indeed, lying has been shown to entail moral costs (Gibson et al., 2013; Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019). In contrast, in the Effort treatment, participants earned \$2 ethically, in exchange of the effort provided in the encoding task. If dishonest individuals subjectively feel entitled to the unethical gain because they incurred a moral cost when lying (Thielmann and Hilbig, 2019), we should observe a similar risk taking behavior by those who earned money unethically in the CEO task and those who earned the same amount through effort provision in the encoding task. This should be the case if both groups exhibit a sense of entitlement.

We formulate our first conjecture as follows:

**Conjecture 1** (Unethical vs. Effort money) *Controlling for their baseline attitudes towards risk, individuals whose previous earnings were earned unethically take similar risk as those who take risk with the money earned ethically from providing effort.*

In the Windfall treatment, participants earned \$2 by luck and without any effort cost. If a moral cost of lying engenders a sense of entitlement, then we anticipate that dishonest individuals should take less risk in the BRET than those whose endowment was a windfall gain from a lottery. We formulate our second conjecture as follows:

**Conjecture 2** (Unethical *vs.* Windfall money) *Controlling for their baseline attitudes towards risk, individuals whose previous earnings were earned unethically take less risk with these earnings than those whose previous earnings were windfall money.*

### 3.5 Results of the Main Experiment

As we are interested in comparing the risk taking of individuals who earned money unethically to those who earned money through luck and effort, we focus on the participants who earned \$2 in part 1 of the experiment. The analysis of participants who earned \$1 is reported in Section 3.D in the Appendix.

Before testing our conjectures formally, we started by examining the determinants of the probability of earning \$2 in each treatment, because the decisions to lie and exert effort are both endogenous. Table 3.B.4 in the Appendix reports, for each treatment, a Probit regression in which the dependent variable is equal to one if the participant earned \$2, and zero otherwise. The independent variables include the main socio-demographics: the mean response to the two SOEP questions, normalized in a z-score, gender, the z-score of age, a dummy for the educational level (equal to one for above high school, and zero otherwise), and the mean weekly expenditures. As expected, no variable is significant in the Windfall treatment. The baseline risk attitudes, gender and expenditures did not affect the probability of belonging to the \$2 group in any treatment. Education impacts the likelihood only in the Effort treatment. The coefficient of age is negative and significant at the 1% level in the Lying and Effort treatments, which indicates the presence of self-selection that we need to take into account in the data analysis.

Regarding behavior in the BRET, on average, participants who earned \$2 collected 23.85 boxes in the BRET in the Effort treatment (SD=18.75), 27.33 boxes in the Lying treatment (SD=21.51), and 28.23 boxes in the Windfall treatment (SD=20.97). Two-sided Mann-Whitney rank-sum tests, with one observation per individual, indicate that the number of boxes collected in the Lying treatment is significantly different neither from the number collected in the Effort treatment ( $p = 0.162$ ), nor from the number collected in the Windfall treatment ( $p = 0.559$ ). The pairwise comparison of the Effort and Windfall treatments shows a significant difference ( $p = 0.051$ ). Figure 3.1 displays the Cumulative Distribution Function of the number of boxes collected in the BRET, by treatment. Kolmogorov-Smirnov tests reveal that the distribution of decisions in the BRET differs significantly between the Lying and the Effort treatments ( $p = 0.046$ ), but not between the Lying

and the Windfall treatments ( $p = 0.979$ ). The pairwise comparison of the Effort and the Windfall treatments also reveals a significant difference ( $p = 0.033$ ).

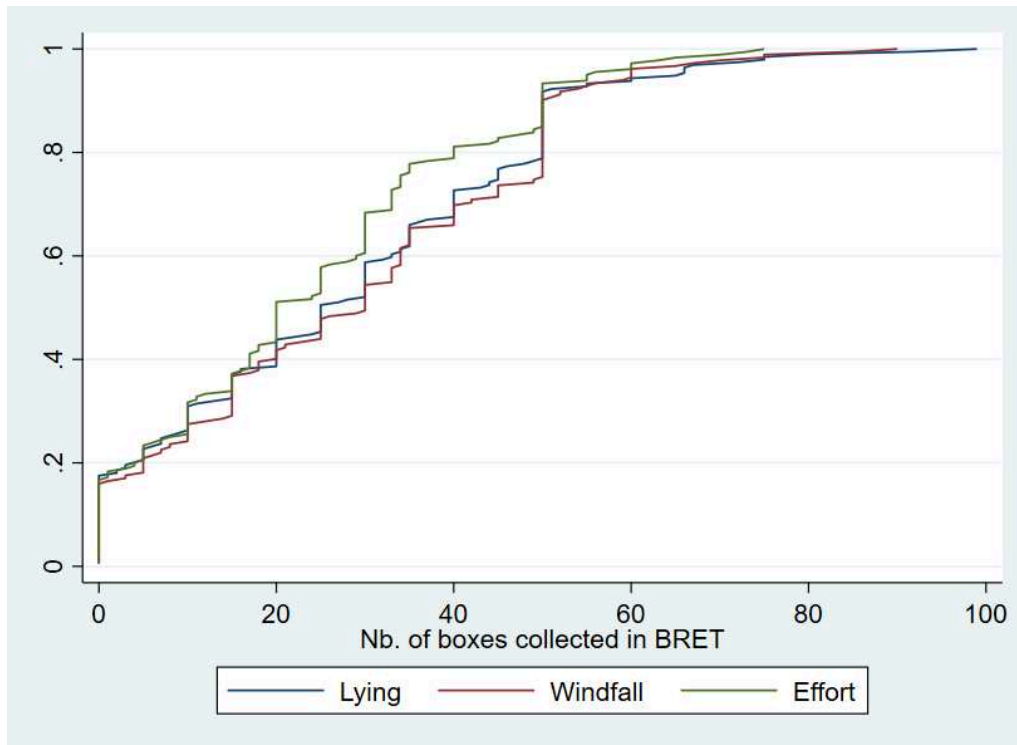


Figure 3.1: CDF of the Number of Boxes Collected in the BRET, by Treatment

To take into account the endogeneity of the inclusion of individuals in the \$2 group in two out of the three treatments, we estimated Poisson regressions with sample selection.<sup>8</sup> In the selection equation, the independent variables include treatment dummies interacted with the baseline risk attitudes (the z-score of the mean response to the two SOEP questions), gender, the z-score of age, education (coded one if the participant's education level is above high school, and zero otherwise), and the mean weekly expenditures. In the outcome equation, the dependent variable is the number of boxes collected in the BRET. In model (1), the independent variables include the treatment dummies, with the Effort treatment as a reference category. The control variables are the z-score of the baseline risk attitudes, a dummy for gender and the z-score of age.<sup>9</sup>

<sup>8</sup>This model handles endogenous sample selection like a Heckman selection model, but it is more appropriate for count data. Using regressions with sample selection instead of OLS is a deviation from the pre-registered data analysis plan.

<sup>9</sup>Exclusion of a significant selection variable from the outcome equation is not necessary, but often recommended in the literature whenever possible (see Heckman, 1978; Wilde, 2000; Miranda and Rabe-Hesketh, 2006). However, as age is an important determinant of risk taking behavior, we did not exclude this variable in the outcome equation, although it is significant in the selection equation. The Information Criteria of the models with age indicate a better fit for our data than models that exclude age from the outcome equation.

Model (2) replicates model (1), but further includes interaction terms between the treatment dummies and the z-score of the mean response to the two SOEP questions in the independent variables. This accounts for the fact that individuals with various baseline risk attitudes might react differently to the nature of the money at stake. Model (2) allows us to capture the marginal effects of the treatments at each given level of baseline risk attitude.

Table 3.1 reports the marginal effects from these two Poisson regressions with sample selection. The estimated correlation between the selection errors and the outcome errors ( $\rho$ ), log pseudo-likelihood and the Akaike's and Schwarz's Bayesian Information Criteria (AIC and BIC) are reported below each model. The Wald tests of independent equations indicate that  $\rho$  is statistically significant ( $p < 0.01$ ), confirming the need for two-stage regressions.

Table 3.1: Determinants of Risk Taking in the BRET in the \$2 Group

Dep. Var.	(1)		(2)	
	ME	St.Err.	ME	St.Err.
Nb. of boxes				
Effort treatment	<i>Ref.</i>		<i>Ref.</i>	
Lying treatment	7.685**	(3.671)	6.194**	(2.426)
Windfall treatment	22.84***	(3.619)	6.961**	(2.716)
Mean SOEP	11.43***	(1.470)	11.93***	(1.240)
Male	3.466	(3.185)	2.930	(2.541)
Age	-9.616***	(1.612)	-6.013***	(1.731)
$N$	1033		1033	
$N$ Selected	556		556	
$\rho$	-0.46		-0.33	
Treatment x SOEP	No		Yes	
Log pseudo-likelihood	-3212.97		-3194.28	
AIC	6477.94		6444.57	
BIC	6606.39		6582.89	

*Notes:* This table reports the average marginal effects (ME) from the outcome equation of Poisson regressions with sample selection into the \$2 group. Delta method standard errors (St. Err.) are in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table 3.1 shows that our experiment replicates the entitlement effect previously identified in the literature: participants who earned money through effort provision were more conservative in their risk taking than those who earned the same amount through luck. It also reveals that, contrary to Conjecture 1, the participants who earned a higher endowment by lying took more risk in the BRET than those who



earned their endowment through effort provision. Model (1) indicates that individuals who earned money through lying or through luck collected 7 to 23 more boxes, respectively, than those who earned the same through effort. The difference between the Windfall and the Effort treatments is significant at the 1% level, and the difference between the Lying and the Effort treatments at the 5% level. The marginal effects of the Lying and the Windfall treatments are significantly different from each other (Wald test,  $p < 0.001$ ).

Model (2), which accounts for the interaction effects of the treatments and the baseline risk attitudes, shows that individuals with average baseline risk attitudes who earned money dishonestly collected 6 more boxes than those in the Effort treatment, and those who earned money through luck collected almost 7 more boxes, both being significantly different from the participants in the Effort treatment at the 5% level. In contrast with model (1), the marginal effects of the Lying and the Windfall treatments are no longer statistically different (Wald test,  $p = 0.809$ ). Based on the values of AIC and BIC, model (2) fits the data better. This suggests that the interaction terms between the treatment dummies and the baseline risk attitudes play an important role. We will return to this point in the next section.

This analysis supports the following results that reject our conjectures.

**Result 1** (Unethical *vs.* Effort money). Individuals who earned money dishonestly took more risk in the BRET than those who earned the same amount of money honestly through providing effort. Conjecture 1 is not supported.

**Result 2** (Unethical *vs.* Windfall money). There is no significant difference in risk taking between individuals who earned money dishonestly and those who earned the same amount of money through luck. Conjecture 2 is not supported.

## 3.6 Follow-up Experiment

Our first two results suggest that liars treated their money more like a windfall gain from luck than like an effort-based earning. However, we cannot exclude that this finding was driven by self-selection: individuals who lied in the CEO task, although the experimenters were *ex-post* able to detect lying at the individual level, were perhaps individuals with a low sensitiveness to morality. If this is true, it might still be the case that a moral cost may generate a feeling of entitlement but the effect was simply not captured in our experiment due to the selection of low moral

cost individuals. To explore this hypothesis, we ran a follow-up experiment with a new treatment that increased the *ex-post* moral cost of lying. We recruited new participants in the United States through MTurk and collected 354 observations in the *Higher Moral Cost (Lying)* treatment.<sup>10</sup>

**Design:** To increase the moral cost of lying in the new treatment while keeping comparability with the original Lying treatment, we informed participants, after they made their decision in the CEO task, about the social norms elicited in an independent study. The norm elicitation study was conducted prior to the Higher Moral Cost treatment with different participants. Fifty individuals located in the United States were recruited through MTurk to participate in this study. After reading the description of the CEO task, they were asked to rate the social appropriateness of the two possible decisions in this task (*i.e.*, telling the truth by announcing 31 and telling a lie by announcing 35), by selecting from a 4-point likert scale (*i.e.*, very socially appropriate, somewhat socially appropriate, somewhat socially inappropriate, and very socially inappropriate). They did not play the game themselves. As in [Krupka and Weber \(2013\)](#), they received a bonus if their response coincided with the modal answer given by all the other participants. One decision was randomly selected for payment.

We then showed the distributions of the ratings of social appropriateness to the participants in the Higher Moral Cost treatment to manipulate the *ex-post* moral cost of lying. Precisely, the procedure was similar to that implemented in the Lying treatment, except that after completing the CEO task, participants received unexpected information about the social norms in this task. They were first informed of the social appropriateness ratings of the truth-telling option, followed by the social inappropriateness ratings of the lying option.<sup>11</sup> The distributions of ratings were presented in a graphical format (see Figures 3.C.1 and 3.C.2 in the Appendix). These figures clearly conveyed that choosing the lying (honest, respectively) option in the CEO task was deemed socially inappropriate (appropriate, respectively) by

---

<sup>10</sup>The number of observations was based on the power calculation we pre-registered. Assuming a type-I error rate of 0.05, a power level of 0.8 and a small effect size (Cohen's  $d = 0.3$ ), the required number of observations to uncover the hypothesized difference between the Higher Moral Cost and the Lying treatments was 184 observations per treatment. However, based on our initial experiment, we conservatively assumed a lying rate of 50% and set the required number of observations for Higher Moral Cost treatment to be 368. We collected 369 observations but we excluded ten individuals who did not pass the attention checks and five individuals who collected 100 boxes, revealing misunderstanding of the instructions.

<sup>11</sup>There was a subtle trade-off here: the objective of the new treatment was to increase the moral cost of lying without raising an ethical issue. This is why we provided information about the two norms.

the majority of people. After receiving information on social norms, participants proceeded to the risk task.

In sum, the new treatment intended to increase the *ex-post* moral cost of lying, while keeping the same selection as in the Lying treatment. If a moral cost generates a feeling of entitlement to the dishonestly earned money, this feeling should be stronger among liars who experienced a higher moral cost than among those who incurred a lower cost. Thus, we conjectured that liars in the Higher Moral Cost treatment should take less risk in the BRET than liars in the Lying treatment.

**Conjecture 3** (Increasing the moral cost of lying and entitlement) *Controlling for their general attitudes towards risk, dishonest individuals who incurred a higher moral cost take less risk than those who incurred a lower moral cost. This should hold if the moral costs of lying induce a feeling of entitlement.*

Alternatively, increasing the moral cost of lying might induce a disentanglement effect, as seen in [Chen et al. \(2017\)](#). This is plausible if liars in the Higher Moral Cost treatment were in a state of high cognitive dissonance (*i.e.*, they felt guilty of what they did in the CEO task after observing the social norms), and thereby, decided to take even higher risk in the BRET as a way to disown the unethical gain or reduce its share in total earnings from the experiment. We thus formulate the following alternative conjecture:<sup>12</sup>

**Conjecture 3A** (Increasing the moral cost of lying and disentanglement) *Controlling for their general attitudes towards risk, dishonest individuals who incurred a higher moral cost take more risk than those who incurred a lower moral cost. This should hold if the moral costs of lying induce a disentanglement effect due to cognitive dissonance.*

Unsurprisingly, the proportion of liars in the Higher Moral Cost treatment (54.2%) is not significantly different from the proportion of liars in the Lying treatment (52.3%) ( $p = 0.600$ , chi squared test). In the new treatment, liars collected on average 27.55 boxes in the BRET (SD= 20.61). The Cumulative Distribution Function of the collected boxes is displayed in Figure 3.C.3 in the Section 3.C in the Appendix.<sup>13</sup> To test Conjecture 3 against Conjecture 3A, we reestimated the same

---

<sup>12</sup>The design and conjectures related to the follow-up experiment were pre-registered separately at AsPredicted (#67936) after we collected the data of the main experiment but before we collected the data of the new treatment.

<sup>13</sup>Two-sided Mann-Whitney rank-sum tests indicate that the number of boxes collected in the

Poisson models with sample selection as in Table 3.1, pooling the data from the main treatments and those from the Higher Moral Cost treatment. The results are reported in Table 3.2.

Table 3.2: Determinants of Risk Taking in the BRET in the \$2 Group (Original and Follow-up Experiments)

Dep. Var.	(1)		(2)	
	ME	St.Err.	ME	St.Err.
Nb. of boxes				
Effort treatment	<i>Ref.</i>		<i>Ref.</i>	
Lying treatment	8.978***	(2.979)	5.683***	(1.848)
Windfall treatment	22.86***	(4.212)	6.776*	(3.615)
Higher MC treatment	0.641	(2.318)	-1.094	(2.524)
Mean SOEP	10.42***	(1.446)	11.41***	(0.934)
Male	1.745	(2.580)	0.826	(3.327)
Age	-7.152***	(1.106)	-4.639***	(1.008)
<i>N</i>	1387		1387	
<i>N</i> Selected	748		748	
$\rho$	-0.40		-0.30	
Treatment X SOEP	No		Yes	
Log pseudo-likelihood	-4331.15		-4303.65	
AIC	8728.29		8679.29	
BIC	8901.04		8867.74	

*Notes:* This table reports the marginal effects (ME) from the outcome equation of Poisson regressions with sample selection into the \$2 group. Delta method standard errors (St. Err.) are in parentheses. The sample includes the data from all the treatments, including the High Moral Cost treatment. \*  $p < 0.10$ , \*\*\*  $p < 0.01$ .

Two points are noteworthy in the regressions reported in Table 3.2. First, confirming the results from Table 3.1, individuals who earned money dishonestly in the Lying treatment or by being lucky in the Windfall treatment took significantly more risk in the BRET than those who earned the same amount through effort (they collected approximately 9 more boxes, according to model (1), and 6 more boxes, according to model (2), in the Lying treatment; the respective numbers are

Higher Moral Cost treatment is not significantly different from the number collected in the other treatments ( $p = 0.962$  in a pairwise comparison with the Lying treatment;  $p = 0.092$  with the Effort treatment;  $p = 0.580$  with the Windfall treatment). Kolmogorov-Smirnov tests also reveal no significant differences in the distribution of decisions in the BRET between the Higher Moral Cost treatment and the Lying treatment ( $p = 0.559$ ), the Effort treatment ( $p = 0.311$ ), and the Windfall treatment ( $p = 0.267$ ).

23 and 7 in the Windfall treatment). Additionally, at average baseline risk attitudes, the initial significant difference between the marginal effects of the Lying and the Windfall treatments (Wald test;  $p < 0.001$  in model (1)) disappears ( $p = 0.762$  in model (2)), suggesting that dishonest and lucky individuals with average baseline risk attitudes took similar levels of risk in the BRET. Second, liars in the Higher Moral Cost treatment did not collect a significantly different number of boxes than participants in the Effort treatment. They collected significantly less boxes than liars in the original Lying treatment (Wald tests,  $p = 0.002$  in model (1);  $p = 0.006$  in model (2)), and than lucky players in the Windfall treatment ( $p < 0.001$  in model (1);  $p = 0.010$  in model (2)). This suggests that the manipulation of the moral cost induced a feeling of entitlement. This analysis supports the following result:

**Result 3** (Increasing the Moral Cost of Lying). Dishonest individuals who were exposed to a higher moral cost took significantly less risk in the BRET than liars whose moral costs were not manipulated. Conjecture 3 is supported against Conjecture 3A.

Finally, the model with interactions in Table 3.2 allows us to examine the heterogeneity of reactions to the nature of money, based on individuals' baseline risk attitudes. Using the estimates from model (2), Figure 3.2 plots the marginal effects of the Lying, Windfall and Higher Moral Cost treatments (the Effort treatment is the reference category) for given baseline risk attitudes.

This figure delivers interesting information. Though the marginal effects of the Lying and Windfall treatments from Table 3.2 are positive and not significantly different from each other, they are in fact driven by different groups of individuals. In the Lying treatment, the treatment effect is largely driven by participants who were risk averse at their baseline, whereas in the Windfall treatment, it is driven by individuals who were risk loving at their baseline. Lastly, when the moral cost of lying was increased, risk taking with dishonestly earned money was reduced, especially for risk averse liars.

## 3.7 Discussion and Conclusion

In this chapter, we investigated whether individuals make different risk decisions according to the source of the money they put at risk, and, specifically, whether they treat dishonest money differently than money earned honestly through either

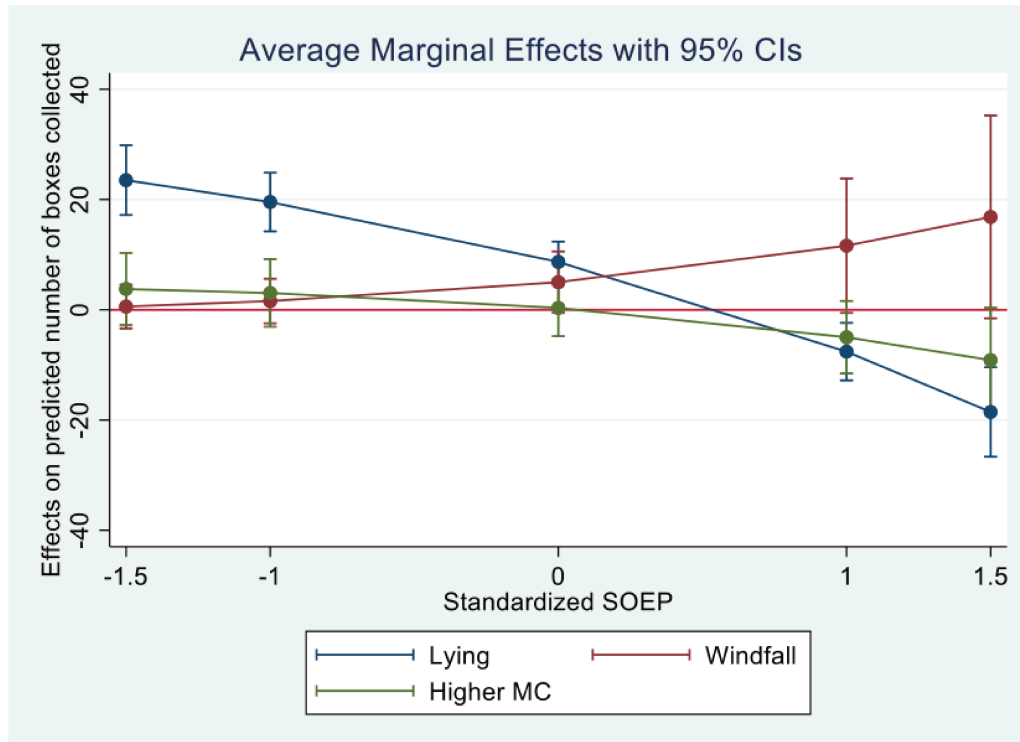


Figure 3.2: Average Marginal Effects of the Treatment Dummies on Risk Taking, Conditional on Baseline Risk Attitudes

*Notes:* This figure displays the marginal effects of each treatment on the number of boxes collected in the BRET (Y-axis), taken from Poisson regressions with sample selection (model (2) in Table 3.2). The horizontal line corresponds to the reference category (Effort treatment). The X-axis represents baseline risk attitudes, as measured by the z-score of the mean responses to the two SOEP questions (zero corresponds to the average risk attitude, that is, slightly risk averse players).

effort or luck. If the moral cost of lying induces a sense of entitlement in the same way as an effort cost (as suggested by Thielmann and Hilbig, 2019), risk taking should be relatively similar when the money put at risk was obtained dishonestly and when it was earned honestly through effort; in contrast, it should be treated more conservatively than when earned through luck.

Contrary to these conjectures, liars took significantly *more* risk with this money in a subsequent task than those who earned the same amount honestly through a cognitive effort. We found no difference in risk taking with the money earned from lying or from luck. This reveals another facet of the non-fungibility of money and suggests that dishonest individuals in our experiment treated unethical money more like a windfall gain than an effort-based one.

However, when increasing the *ex-post* moral cost of lying, via social norms information on the (in)appropriateness of (mis-)reporting, we found that liars took significantly less risk in the BRET than the other liars. This suggests that a higher

moral cost of lying can induce a feeling of entitlement to the dishonestly earned money, resulting in more conservative risk behavior. We also found that the increase in the willingness to take risk with unethical money earned at a lower moral cost came from risk averse individuals. This effect vanished when the moral cost of lying was increased.

Regarding the mechanisms driving our findings, laundering money is unlikely to be the motivation behind increased risk taking with money earned by lying in our study. In one of their experiments [Imas et al. \(2020\)](#) found that liars were more willing to expose larger sums of unethical money to risk when the money would be laundered than when it would not be laundered. The premise of mental money laundering in [Imas et al. \(2020\)](#) rests on the notion of a physical exchange of cash for the same amount from a different source. This feature is not present in our setting. In addition, their lottery task was designed to launder money, with a negligible risk of losing money in the lottery. In contrast, our task involved deciding how much risk to take, knowing that gains would be added to the previous unethical earnings and that part of the money earned previously might be lost (*i.e.*, it would never be exchanged for “clean” money). Pooling “dirty” and “clean” money could possibly be like laundering money, but in our context collecting more boxes in the BRET also meant increasing the risk of losing this money. Moreover, we found that when the moral cost of lying was increased, liars took less risk with this money, while we would expect the opposite if higher risk taking was motivated by the willingness to launder money by pooling the two types of money.

We can also rule out that guilt is the mechanism driving our findings. Indeed, if guilt was the motivation behind higher risk taking in the Lying treatment than in the Effort treatment (for example, in an effort to distancing oneself from dishonest money), we should have observed even higher risk taking when we increased the moral cost of lying in our follow-up experiment. The data point to the opposite direction.

If the mechanism behind our findings is instead the presence of an entitlement effect induced by the moral cost of lying, why did we not observe it in our main experiment? One potential reason could be that lying in the CEO task does not entail any moral cost. We immediately rule out this possibility because not all players lied in the experiment of [Gibson et al. \(2013\)](#), which, on the contrary, revealed the heterogeneity of moral self-concepts, and because there is evidence that lying in the CEO task is perceived as socially inappropriate ([Huber and Huber, 2020](#)). We argue that the moral cost must in fact be high in this task because lying was observable by the experimenter. This might have created a selection effect whereby only individuals with a low moral cost did lie. Therefore, it is plausible that the

liars did not suffer *enough* moral cost to induce a sense of entitlement. They may have perceived the unethical gain simply as ‘*house money*’ (Thaler and Johnson, 1990). In contrast, when we increased the moral cost of lying *ex-post*, keeping the selection process constant, the willingness to take more risk with unethical money than with effort money disappeared. This supports the notion that moral costs can induce a feeling to entitlement to the dishonestly earned money, but this feeling is sensitive to the level of these costs.

This chapter has also revealed the heterogeneity of the treatment effects depending on the individuals’ baseline risk attitudes. Two particular observations are striking. First, in the Lying treatment individuals who were risk averse at their baseline took more risk in the BRET than the other participants, while their counterparts in the other treatments remained consistent with their baseline preferences. A possible interpretation is that risk averse liars were particularly sensitive to the perception of unethical money as a house money that could be put at stake, as if it had little value, whereas the same individuals would be more conservative with other sources of earnings. Second, on the opposite side of the spectrum, risk lovers who won a lottery in the Windfall treatment took more risk in the BRET than risk lovers in the other treatments. This might result from a hot-hand effect induced by the gambling environment, whereas in the other treatments the money that could be put at stake had to be earned. These findings suggest that studies of the violations of the fungibility of money would benefit from a systematic attention to the individuals’ risk preferences.

Overall, Chapter 3 provides evidence of another type of violation of the fungibility of money by showing that the mere nature of unethical gains matters for risk taking. Although such violation is not morally motivated in our settings, it shows that the treatment of unethical money is sensitive to the level of the moral cost of dishonesty. Earning unethical money at a higher moral cost led to a more conservative use of this money, as if it was hard-earned, whereas individuals –especially risk averse ones– treated unethical money at a low moral cost more like a windfall gain from luck. We need to remain cautious about the external validity of these findings in a more complex environment. Nevertheless, they could contribute to the reflection on risk taking in domains such as financial markets in which the exploitation of asymmetric information may be tempting. While easy money encourages risk taking, making the inappropriateness of earning money unethically more salient could induce a more conservative use of money, and perhaps limit excessive gambling and misallocation of clients’ portfolio (Foerster et al., 2017; Kling et al., 2019).



## Appendix

### 3.A Instructions

#### *Lying treatment*

#### Participant information statement

**1. What does the study involve?**

This study involves several tasks and a short questionnaire. We *strongly recommend* that you complete this study using a computer/laptop or a tablet/ipad.

**2. Who is carrying out the study?**

The study is being conducted by Professor Marie Claire Villeval and Sorravich Kingsuwanukul from the University of Lyon.

**3. How much time does it take?**

This study should take approximately **10 to 15 minutes** to complete.

**4. Can I withdraw from the study?**

Participation is voluntary. If you do consent to participate, you can still withdraw from the study at any time without penalty and without having to give any reason. However, there will be no payment in case you withdraw. Withdrawing will not affect the relationship between you, the researchers and Mturk in any way. Once you have completed the study anonymously, your responses cannot be withdrawn.

**5. Will anyone else know the results?**

All aspects of the study, including results, will be **confidential and anonymous**. We do **NOT** ask for your name or any information that will make you identifiable. A report of the study may be submitted for publication, but all information will be used in an aggregated form, and you will never be identified.

**6. Will this study benefit me?**

By participating in this study, you will receive a flat payment of \$0.50, **plus** an additional payment. This additional payment depends on the decisions you make in the study.

**7. Can I tell other people about the study?**

The researchers request that, for the purpose of maintaining study integrity, you do not share with anybody the nature of the questions asked of you.

**8. What if I require further information about the study?**

If you have any specific question regarding the study, please contact us via email at [kingsuwanukul\[at\]gate.cnrs.fr](mailto:kingsuwanukul[at]gate.cnrs.fr).

**If you agree to the above conditions, please select 'I agree' and then click 'Next' to proceed.**

I agree  I do not agree

NEXT

## Overview

You are now participating in the study. It should take approximately **10 to 15 minutes** to complete. We request you not to stop in between. After you finish the study, you will receive a flat payment of \$0.50, **plus an additional bonus** depending on the decisions you make during the study within 48 hours. **The exact amount depends on your decisions, so please pay attention.**

Your job today is to perform two tasks and participate in a brief questionnaire.

Please, press 'Next' to continue.

NEXT

## Questions

Before we proceed, please respond to the questions:

Q1 How do you see yourself: are you generally a person who is fully prepared to take risk or do you try to avoid taking risks?

Please give a value between 0 and 10, with 0 for "not at all willing to take risk" and 10 for "very willing to take risks":

0    1    2    3    4    5    6    7    8    9    10

Q2 How would you rate your willingness to take risks concerning financial matters?

0    1    2    3    4    5    6    7    8    9    10

Please, press 'Next' to continue.

NEXT

### First task

The first task is called 'CEO Task'

Imagine you are the Chief Executive Officer (CEO) of a publicly listed company and one of your tasks is to inform shareholders each quarter about the course of business and the earnings per share. There are two options: announcing 31 cents per share or 35 cents per share. **The higher the announced earnings, the higher will be your compensation as the CEO.** As the CEO, you know the actual earnings (31 cents per share), but shareholders do not. Shareholders anticipate the announcement of 35 cents as earnings per share, but you know that earnings of 31 cents per share would more accurately reflect the actual number.

The earnings you announce determine how much money you will be paid for this task.

- Announcing 31 cents per share when the actual earnings per share is 31 cents. In this case, you will earn \$1.
- Announcing 35 cents per share when the actual earnings per share is 31 cents. In this case, you will earn \$2.

Please, press 'Next' to continue.

NEXT

### First task

In the situation described, how many cents of earnings per share will you announce?

- Announce 31 cents per share when the actual earnings per share is 31 cents. In this case, you will earn \$1.
- Announce 35 cents per share when the actual earnings per share is 31 cents. In this case, you will earn \$2.

After making your choice, please, press 'Next' to continue.

NEXT

### End of the first task

You completed the CEO task. You chose to announce **31** cents per share. Therefore, your earning for this task is **\$1**.

Please, press 'Next' to continue.

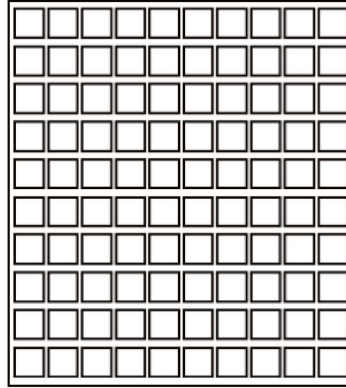
NEXT

## Second task

The second task is called "Grid Task".

You have earned \$1.00 in the CEO task. In this second task, \$0.50 of this earning is put at stake.

You will see a grid representing 100 boxes, as shown below. The boxes are numbered from 1-100. Starting from the top-left corner.



The value of each box is \$0.03. Your task is to choose how many boxes to collect. So, you will be asked to choose a number between 0 to 100. Boxes will be collected in numerical order.

99 of these boxes are empty, but the program has hidden a 'bomb' in one box. You do not know where the bomb lies. You only know that the bomb can be in any box with an equal chance. Moreover, even if you collect the bomb, you will not know it until the end of the study.

At the end of the study, the program will randomly determine which box contains the bomb. Your earning for this task depends on whether you have collected the box that contains the bomb or not.


Please, press 'Next' to continue.

NEXT

## Second task

There are 2 possible situations.

- If you have only collected empty boxes, you will earn \$0.03 for each box collected.
- If you have collected the box that contains the bomb, you will earn nothing from collecting the boxes and \$0.50 will be deducted from what you earned in the CEO task (\$1.00).

When you choose a number of boxes you would like to collect, you have to write that number twice. Then, the program will display a grid with collected boxes marked . Once you are satisfied with your decision, you have to click 'Confirm' to submit your choice.

**Note:** If you do not wish to lose any of your earning from the CEO task, you can choose to collect zero box.

Before the actual task, you will have to respond to some questions and perform a practice round. **This gives you the opportunity to understand the task and how your decision may affect your earnings.** The decision in the practice round does **NOT** count towards your additional payment. After the practice round, the actual task will begin.

When you are ready, please, press 'Next' to proceed to the quiz.

NEXT

## Questions

Before we proceed to the practice round, please answer the following question.

1. When will you know about the location of the bomb?

- Immediately when you collect the box that contains a bomb
- At the end of the study

NEXT

## Questions

Before we proceed to the practice round, please answer the following question.

2. What happens if one of the boxes you collected contains the bomb?

- Nothing happens
- You lose all the earnings from the collected boxes
- You lose all the earnings from the collected boxes and \$0.5 is deducted from your earning from the CEO task

NEXT

## Questions

Before we proceed to the practice round, please answer the following question.

3. Jane decided to collect 25 boxes. What is the chance that Jane find a bomb?

- 75 chances out of 100
- 25 chances out of 100
- 100 chances out of 100

NEXT

## Questions

Before we proceed to the practice round, please answer the following question.

4. John decided to collect 25 boxes. Therefore, the program collected boxes no. 1 to 25 for him. At the end of the study, the program randomly determined the location of the bomb, among 100 boxes, and it is behind box no. 52. Did John collect the bomb?

- Maybe
- Yes
- No

NEXT

## Practice round

You have completed the quiz about the Grid Task. The next page is a **practice**, so you can get to know the interface of the task. **Your decision in this practice round is NOT counted for any payment.**

When you are ready, please, press 'Next' to begin.

NEXT

## Practice round

Your decision in this practice round is **NOT** counted for any payment.


Please enter the number of boxes you like to collect

Please enter the same number again

Then, please click 'Confirm my decision' to continue.

**Confirm my decision**

Practice round

Your decision in this practice round is **NOT** counted for any payment.

✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Please enter the number of boxes you like to collect

Please enter the same number again

Then, please click 'Confirm my decision' to continue.

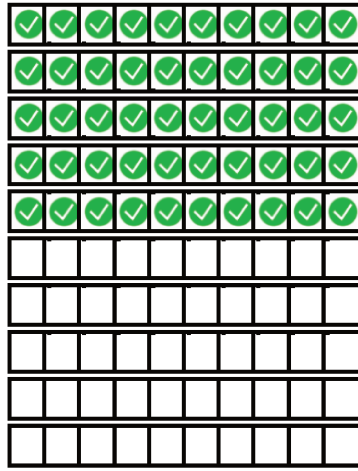
Confirm my decision

## Practice round

Your decision in this practice round is **NOT** counted for any payment.

You decided to collect **50** boxes.

- If there is no bomb behind the selected boxes, you would earn **\$1.5** for this task.
- If there is a bomb behind the selected boxes, you would earn **nothing** from collecting the boxes and **\$0.50 would be deducted** from your earning in the CEO task (\$1.00).



Please, click 'Next' to continue.

NEXT

## End of Practice round

This is the end of the instruction for the Grid Task

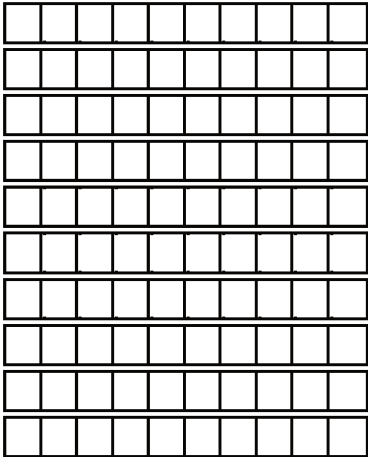
On the next page, you will find the actual Grid Task.

When you are ready, please, press "Next" to begin.

NEXT



### Grid Task



Please enter the number of boxes you like to collect

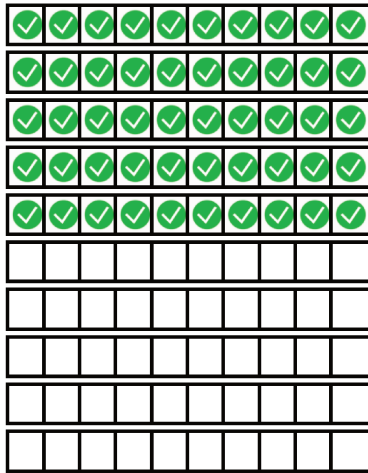
Please enter the same number again

Then, please click 'Confirm my decision' to continue.

**Confirm my decision**

### Grid Task



Please enter the number of boxes you like to collect

Please enter the same number again

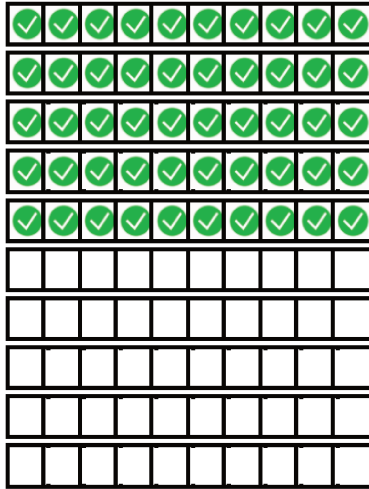
Then, please click 'Confirm my decision' to continue.

**Confirm my decision**

## Grid Task

You decided to collect **50** boxes.

- If there is no bomb behind the selected boxes, you earn **\$1.5** for this task.
- If there is a bomb behind the selected boxes, you earn **nothing** from collecting the boxes and **\$0.50 is deducted** from your earning in the CEO task (\$1.00).



Please, click 'Next' to continue.

NEXT

## The end of second task

You have completed the Grid task.



Please, press "Next" to continue.

NEXT

## Feedback of the bomb location

You have completed all the tasks in this study!

The program will now randomly determine which box contains a bomb. Each box is equally likely to contain the bomb.

- If you collected the bomb, a red bomb  will be shown.
- If you did **NOT** collect the bomb, a black bomb  will be shown.

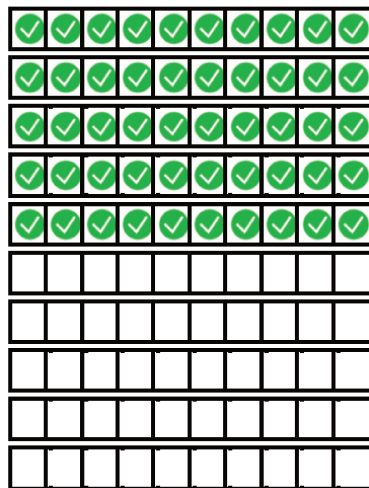
Then, you will answer a brief questionnaire. At the end, the program will summarize your compensation for today's study.

Please, press 'Next' to continue.

NEXT

## Your decision in the Grid Task

You decided to collect **50** boxes.



Please, click 'Reveal bomb' to continue.

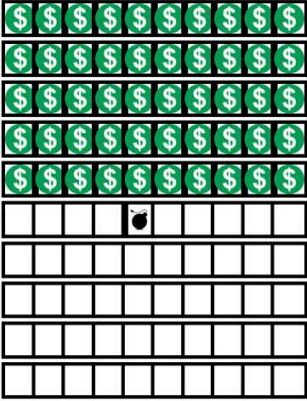
Reveal bomb

After clicking 'Reveal Bomb' and the bomb was not collected ...

**Your decision in the Grid Task**

You decided to collect 50 boxes. The bomb was behind box number 55.

You did not collect a bomb. Therefore, you earned \$1.5 for this task.



Please, click 'Next' to continue.

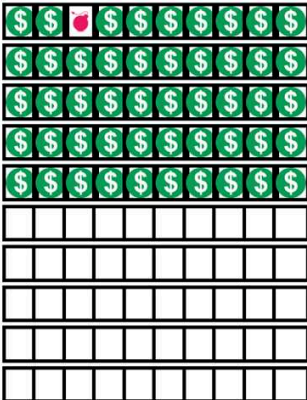
NEXT

After 'Reveal Bomb' and the bomb was collected ...

**Your decision in the Grid Task**

You decided to collect 50 boxes. The bomb was behind box number 3.

You collected a bomb. Therefore, \$0.50 is deducted from your earning in the CEO task.



Please, click 'Next' to continue.

NEXT

## Final questionnaire

Please provide the following information.

1. What year were you born?

2. What is your gender?

- Male
- Female

3. What is your status?

- Student
- Unemployed
- Employed or self-employed
- Pensioner
- Other

4. What is the main source of personal finance?

- Allowance and/or social benefits
- Salary
- Pension

5. What is the size of your household?

- I live alone
- Two persons
- Three persons
- Four persons
- Five persons
- More than 5 persons

6. What is the highest education degree obtained?

- Below high school
- High school
- Bachelor / College degree
- Master / graduate degree and higher

Final Survey (*Continued*)

7. What is your approximate household annual pre-tax income?

- Less than \$10,000
- Between \$10,000 and \$20,000
- Between \$20,000 and \$30,000
- Between \$30,000 and \$50,000
- Between \$50,000 and \$70,000
- Between \$70,000 and \$90,000
- Between \$90,000 and \$110,000
- Between \$110,000 and \$130,000
- Between \$130,000 and \$150,000
- More than \$150,000

8. How much money do you spend in a typical week? This should include your daily expenditure such as food, travel, mobile charges, purchases, **EXCLUDING** rent, mortgage, educational fee, work expenses.

- \$0 - \$29
- \$30 - \$49
- \$50 - \$79
- \$80 - \$119
- \$120 - \$174
- \$175 - \$249
- \$250 - \$350
- More than \$350

9. What device are you using to respond to this study?

- Computer/laptop
- Tablet/Ipod
- Smartphone

10. How easy was it for you to understand the descriptions of the tasks in this study?

Please give a value between 0 and 9, with 0 for "not confusing & easy to understand" and 9 for "confusing & hard to understand":

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

11. What will you get as additional payment today?

- Chocolate
- Gift Card
- Money

12. In what currency will you be paid for this study?

- Swiss Franc
- US Dollar
- Thai Baht

---

NEXT

## The end of today's study

**Thank you! You have completed today's study.**

Here is the summary of your total compensation for participating in today's study.

- Flat payment for participation: \$0.50
- Your earnings in the CEO Task: \$1.00
- Your earnings in the Grid Task: \$1.50

Thus, a total of **\$3.00** will be posted in your MTurk Account within 48 hours.

Please, press "Next" to continue.

NEXT

## The end of the study

Thank you for taking time out of your busy life to participate in this study. If you have any questions concerning this study, you can contact us at [kingsuwankul\[at\]gate.cnrs.fr](mailto:kingsuwankul[at]gate.cnrs.fr).

Your confirmation code to be entered on Mturk webpage is your Mturk worker ID. Please make the HIT on Mturk with this ID.

You can close this window now.

### *Windfall Treatment*

*(Instructions reported only for part 1 since the other parts were similar to that in the previous treatment.)*

#### Information

You will be receiving an amount of money as an endowment. You are receiving either \$1 or \$2 with an equal chance. The program will determine the exact amount you will receive, which does not depend on your decisions.

When you press 'NEXT', the program will determine the exact amount of your endowment.

NEXT

The program has determined the amount for you

Your endowment is \$1.

Please, press 'Next' to continue.

NEXT



### Effort Treatment

(Instructions reported only for part 1 since the other parts were similar to that in the previous treatment.)

## First task

The first task is an encoding task.

You will be presented with a number of words and your task will be to encode these words by substituting the letters of the alphabet with numbers using the following table:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

**Example:** You are given the word **FLAT**. The letters in the table above show that F=6, L=3, A=8, and T=19. You will have to enter these numbers into boxes corresponding to the respective letters of the word.

Once you encode a word correctly, the program will prompt you with another word to encode. You will have **8 minutes** to encode as many words as you want. After 8 minutes, the task is ended automatically.

**Your earnings for this task:**

- If you correctly encode fewer than 39 words, you will earn \$1.
- If you correctly encode 39 words or more, you will earn \$2.

When you are ready, please press "Next" to start the task.

NEXT

Remaining time:421 seconds

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
8	12	14	10	9	6	24	22	7	5	11	3	18	1	21	16	23	2	13	19	25	4	26	17	20	15

So far, you have encoded 5 words correctly.

C	O	P	Y
<input style="width: 30px; height: 20px;" type="text"/>	<input style="width: 30px; height: 20px;" type="text"/>	<input style="width: 30px; height: 20px;" type="text"/>	<input style="width: 30px; height: 20px;" type="text"/>

NEXT

You will earn \$2 for this task if you encode at least 39 words. Otherwise, you will earn \$1.

### Higher Moral Cost Treatment

(Instructions reported only for the differences with those used for the Lying treatment, that take place between the end of the CEO task and the BRET).)

Before we proceed, please read the following message carefully:

In another study, we asked participants, who are also MTurkers in the U.S. just like you, whether they believed that the participants' decisions to announce an earning of 31 and 35 cents per share when knowing that the actual earning was 31 cents are, according to the majority of the participants, very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, very socially appropriate.

By socially inappropriate, we mean the behavior that most people would agree is the "incorrect" or "unethical" thing to do. By socially appropriate, we mean the behavior that most people would agree is the "correct" or "ethical" thing to do.

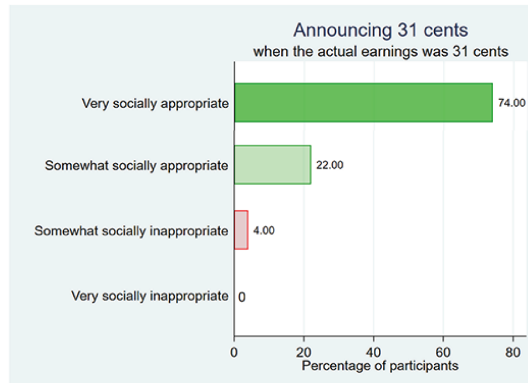
Please, press 'Next' to continue.

NEXT

Please read the information

About announcing 31 cents per share when knowing that the actual earning per share was 31 cents

In the picture below, you can see the percentages of participants who think that announcing 31 cents per share when knowing that the actual earning per share was 31 cents is, according to the majority of the participants, very socially appropriate, somewhat socially appropriate, somewhat socially inappropriate, very socially inappropriate.



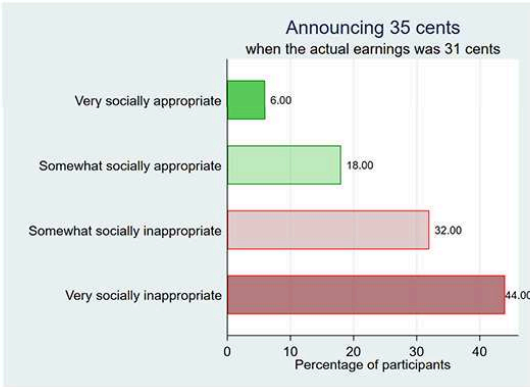
When you finish reading, please click 'Next' to proceed.

Please wait (1)

Please read the information

About announcing 35 cents per share when knowing that the actual earning per share was 31 cents

In the picture below, you can see the percentages of participants who think that announcing 35 cents per share when knowing that the actual earning per share was 31 cents is, according to the majority of the participants, very socially appropriate, somewhat socially appropriate, somewhat socially inappropriate, very socially inappropriate.



When you finish reading, please click 'Next' to proceed.

Please wait (14)

### 3.B Appendix Tables

Table 3.B.1: Summary Statistics (All participants)

	(1)		(2)		(3)		(4)		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)	
	Lying		Windfall		Effort		Higher MC								<i>p</i> -value
	Mean	SD	Mean	SD	Mean	SD	Mean	SD							
Male (Dummy)	0.51	0.50	0.53	0.50	0.50	0.50	0.40	0.49	0.572	0.755	0.003***	0.398	0.001***	0.014**	
Age (Years)	41.52	12.70	40.58	12.88	42.47	12.87	40.61	12.76	0.213	0.297	0.334	0.043**	0.845	0.055*	
Weekly expenditures (1-8)	4.72	1.97	4.57	1.95	4.77	1.90	4.62	1.81	0.219	0.834	0.392	0.159	0.585	0.314	
Above high school (%)	0.72	0.45	0.68	0.47	0.69	0.46	0.71	0.45	0.291	0.479	0.882	0.770	0.371	0.576	
SOEP General (0 - 10)	4.48	2.55	4.55	2.70	4.37	2.46	4.75	2.61	0.697	0.683	0.133	0.442	0.301	0.066*	
SOEP Finance (0 - 10)	4.11	2.64	4.16	2.64	3.97	2.48	4.42	2.71	0.786	0.633	0.125	0.466	0.210	0.046**	
Understanding (0 - 9)	2.41	3.50	2.70	3.53	2.56	3.53	2.49	3.40	0.155	0.388	0.191	0.620	0.841	0.724	
Proportion of \$2	0.52	0.50	0.50	0.50	0.60	0.49	0.54	0.50	0.534	0.036**	0.600	0.007***	0.256	0.113	
Observations	371		364		298		354								

*Notes:* ‘Understanding’ is the mean rating given at the end of the session to the question: ‘How easy was it for you to understand the descriptions of the tasks in this study?’ (0 for “not confusing/easy to understand” and 9 for “confusing/hard to understand”). The *p*-values reported are from chi-square tests for binary variables and rank-sum tests for interval variables. MC for moral cost.

Table 3.B.2: Summary Statistics (\$2 group)

	(1)		(2)		(3)		(4)		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)	
	Lying		Windfall		Effort		Higher MC								<i>p</i> -value
	Mean	SD	Mean	SD	Mean	SD	Mean	SD							
Male (Dummy)	0.53	0.50	0.54	0.50	0.49	0.50	0.41	0.49	0.800	0.416	0.014**	0.295	0.008***	0.109	
Age (Years)	38.80	10.92	39.75	12.65	40.02	11.73	38.52	11.52	0.770	0.345	0.811	0.633	0.595	0.232	
Weekly expenditures (1-8)	4.77	1.94	4.54	1.94	4.79	1.86	4.60	1.79	0.157	0.907	0.291	0.190	0.586	0.379	
Above high school (%)	0.70	0.46	0.71	0.46	0.74	0.44	0.73	0.45	0.784	0.296	0.470	0.447	0.661	0.738	
SOEP General (0 - 10)	4.47	2.55	4.60	2.72	4.36	2.28	4.80	2.67	0.668	0.773	0.188	0.480	0.448	0.093*	
SOEP Finance (0 - 10)	4.10	2.58	4.18	2.64	4.06	2.33	4.47	2.79	0.816	0.935	0.220	0.882	0.351	0.208	
Understanding (0 - 9)	2.34	3.52	2.78	3.56	2.04	3.24	2.54	3.42	0.113	0.713	0.158	0.042**	0.773	0.069*	
Observations	194		182		180		192								

*Notes:* ‘Understanding’ is the mean rating given at the end of the session to the question: ‘How easy was it for you to understand the descriptions of the tasks in this study?’ (0 for “not confusing/easy to understand” and 9 for “confusing/hard to understand”). The *p*-values reported are from chi-square tests for binary variables and rank-sum tests for interval variables. MC for moral cost.

Table 3.B.3: Summary Statistics (\$1 group)

	(1)		(2)		(3)		(4)		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)	
	Lying		Windfall		Effort		Higher MC								<i>p</i> -value
	Mean	SD	Mean	SD	Mean	SD	Mean	SD							
Male (Dummy)	0.49	0.50	0.52	0.50	0.52	0.50	0.40	0.49	0.564	0.669	0.095*	0.932	0.025**	0.055*	
Age (Years)	44.51	13.83	41.41	13.08	46.22	13.65	43.09	13.71	0.025**	0.225	0.370	0.002***	0.268	0.050**	
Weekly expenditures (1 - 8)	4.66	2.00	4.60	1.97	4.75	1.96	4.64	1.84	0.739	0.714	0.887	0.487	0.808	0.605	
Above high school (%)	0.75	0.44	0.66	0.48	0.62	0.49	0.70	0.46	0.073*	0.020**	0.322	0.472	0.449	0.168	
SOEP General (0 - 10)	4.49	2.56	4.51	2.68	4.37	2.71	4.69	2.53	0.881	0.816	0.443	0.729	0.537	0.365	
SOEP Finance (0 - 10)	4.12	2.71	4.14	2.65	3.84	2.70	4.37	2.61	0.897	0.395	0.367	0.331	0.429	0.094*	
Understanding (0 - 9)	2.48	3.50	2.61	3.51	3.34	3.82	2.43	3.37	0.674	0.042**	0.682	0.090*	0.976	0.087*	
Observations	177		182		118		162								

*Notes:* ‘Understanding’ is the mean rating given at the end of the session to the question: ‘How easy was it for you to understand the descriptions of the tasks in this study?’ (0 for “not confusing/easy to understand” and 9 for “confusing/hard to understand”). The *p*-values reported are from chi-square tests for binary variables and rank-sum tests for interval variables. MC for moral cost.

Table 3.B.4: Determinants of the Probability of Earning \$2 in Part 1, by Treatment

Dep. Var.	(1)	(2)	(3)	(4)
Earning \$2 in Part 1	Lying	Windfall	Effort	Higher MC
Mean SOEP	-0.055 (0.071)	0.009 (0.065)	-0.002 (0.085)	0.004 (0.069)
Male	0.043 (0.136)	0.039 (0.135)	-0.202 (0.159)	-0.038 (0.144)
Age	-0.320*** (0.072)	-0.077 (0.067)	-0.328*** (0.077)	-0.227*** (0.069)
Above high school	-0.205 (0.153)	0.151 (0.143)	0.419** (0.166)	0.080 (0.153)
Weekly expenditures	0.001 (0.001)	<-0.001 (0.001)	<0.001 (0.001)	<-0.001 (0.001)
Constant	0.072 (0.171)	-0.108 (0.162)	0.073 (0.185)	0.066 (0.159)
<i>N</i>	371		364	298
				354

*Notes:* The table reports the coefficients from Probit regressions. Standard errors in parentheses. Mean of responses in the SOEP questionnaire and age are reported in z-scores. MC for moral cost. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 3.C Appendix Figures

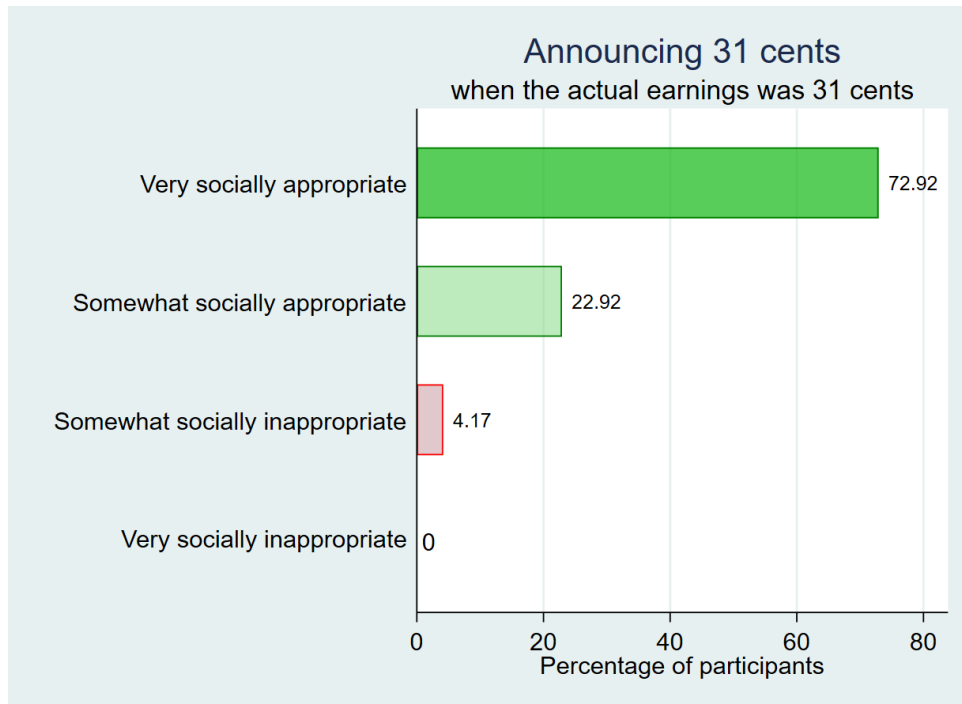


Figure 3.C.1: Higher Moral Cost Treatment: Information on the Social Norm Regarding the Honest Option

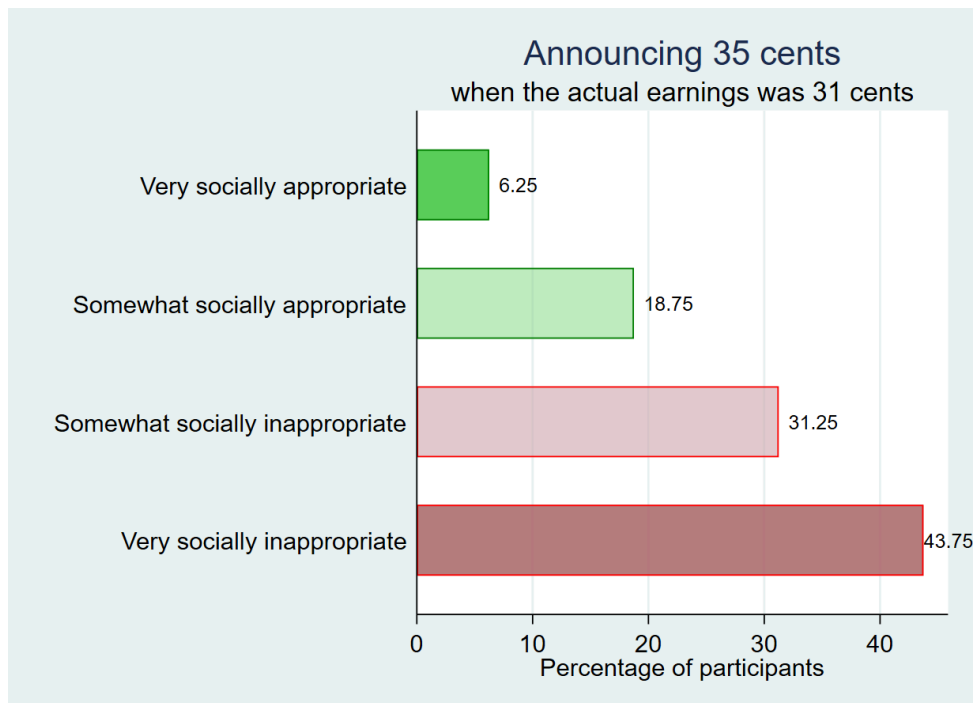


Figure 3.C.2: Higher Moral Cost Treatment: Information on the Social Norm Regarding the Dishonest Option

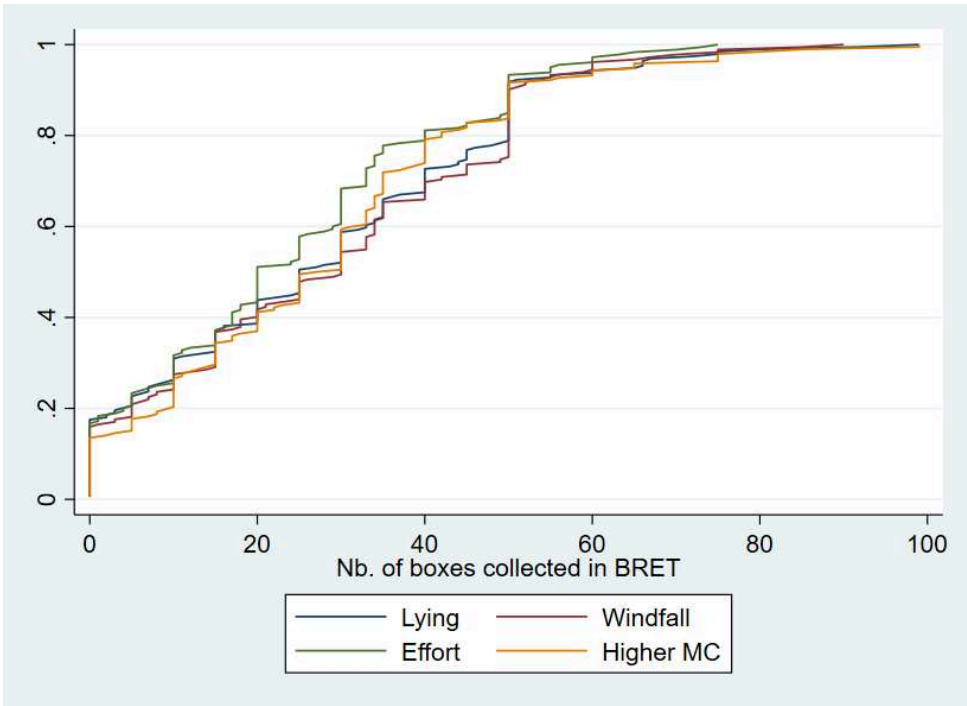


Figure 3.C.3: Cumulative Distribution Function of the number of boxes collected in the BRET, by treatment (including the Higher Moral Cost treatment)

### 3.D Analysis of the \$1 Group Behavior

In this section, we report an exploratory analysis of risk taking behavior in the \$1 group. How participants earned \$1 differed depending on the treatment assigned. In the Windfall treatment, participants received \$1 from the lottery with an equal chance. Therefore, they received \$1 by being unlucky. In the Effort treatment, participants who received only \$1 (which was the base pay irrespective of their performance) are those who fell short of the threshold in the Encoding task (39 words). In the Lying and the Higher Moral Cost treatments, participants who earned \$1 are those who purposefully chose an honest option in the CEO task. Table 3.B.3 summarizes the individual characteristics of these participants.

We performed similar regression analyses as done for the \$2 group. Table 3.D.1 reports the marginal effects from the Poisson regressions with sample selection, in which the dependent variable is the number of boxes collected in the BRET. The selection equation includes interaction terms between each treatment and the baseline risk attitudes (measured as the z-score of the mean response to the two SOEP questions), gender (coded one for male, zero for female), the z-score of age, education (coded one if above high school, zero otherwise), and mean weekly expenditures. In model (1) of the outcome equation reported in Table 3.D.1, the independent variables include the treatment dummies (with the Effort treatment as a reference category), the baseline risk attitudes, gender, and age. Model (2) replicates model (1), but further includes the interaction terms of treatment dummies and the baseline risk attitudes. The estimated correlation between the selection errors and the outcome errors ( $\rho$ ), log pseudo-likelihood, and the Akaike's and Schwarz's Bayesian information criteria (AIC and BIC) are reported for each model.

Table 3.D.1: Determinants of Risk Taking in the BRET in the \$1 group

	(1)		(2)	
Dep. Var.				
Nb. of boxes	ME	St.Err.	ME	St.Err.
Effort	<i>Ref.</i>		<i>Ref.</i>	
Lying	7.588*	(4.285)	8.410***	(1.434)
Windfall	28.71***	(5.206)	18.62***	(2.207)
Higher MC	14.87***	(4.922)	12.16***	(2.470)
Mean SOEP	5.532***	(1.079)	8.802***	(0.634)
Male	0.823	(3.595)	-2.647	(1.756)
Age	-5.345***	(1.762)	-1.069*	(0.647)
$N$	1387		1387	
$N$ Selected	639		639	
$\rho$	-0.25		-0.19	
Treatment X SOEP	No		Yes	
Log pseudo-likelihood	-3820.03		-3802.81	
AIC	7706.07		7677.62	
BIC	7878.82		7866.07	

*Notes:* The table reports the marginal effects from Poisson regressions with sample selection. Delta method standard errors are in parentheses. Mean of responses in the SOEP questionnaire and age are reported in z-scores. \*  $p < 0.10$ , \*\*\*  $p < 0.01$ .

The Wald tests of independent equations indicate that  $\rho$  is statistically significant (both mod-



els,  $p < 0.001$ ). Model (1) indicates that honest and unlucky individuals collected approximately 7 and 28 boxes more than the low performers, respectively. The difference between the marginal effects of the Lying and Windfall treatments is significant (Wald test,  $p < 0.001$ ). Honest individuals in the Higher Moral Cost treatment collected approximately 15 more boxes than low performers, more than the honest individuals in the Lying treatment. The difference between the marginal effects of the Lying and Higher Moral Cost treatments is significant (Wald test,  $p = 0.050$ ). This higher risk taking of honest players in the Higher Moral Cost treatment compared to their counterparts in the Lying treatment should result from the manipulation which made the honesty norm salient and reminded them that others approved of their decision in the CEO task. While this manipulation had the opposite effect on dishonest players, this may have given these honest players the willingness to compensate for the forgone dishonest money in the CEO task by gambling a higher amount in the BRET. Model (2), which includes the interaction terms, yields similar results to those of model (1), with smaller differences across treatments.<sup>14</sup> Our interpretation for the behavior of the \$1 group is that these individuals took risk in the BRET to compensate for having forgone some financial gain prior to the BRET. We discuss these findings below.

**Gambling for resurrection in the Windfall treatment:** We argue that participants who earned \$1 in the Windfall treatment compensated for having been ‘unlucky’ in the lottery in part 1. This interpretation builds on the fact that unlucky participants ( $\text{Mean}_{\text{BRET}} = 31.97$ ,  $\text{SD} = 18.94$ ) tended to collect more boxes than the lucky participants ( $\text{Mean}_{\text{BRET}} = 28.23$ ,  $\text{SD} = 20.97$ ). Though a Mann-Whitney U test does not confirm the significant difference ( $p = 0.097$ , M-W test), a Kolmogorov-Smirnov test indicates that the distributions are significantly different ( $p = 0.002$ ). Table 3.D.2 reports the marginal effects from the Poisson regression for participants in the Windfall treatment, in which the dependent variable is the number of boxes collected in the BRET. The independent variables are the same as in the previous regression tables. The findings in Table 3.D.2 are in line with the gambling for resurrection, as unlucky individuals collected approximately 4 more boxes than lucky ones.

**Entitlement effect in the Effort task:** In the Encoding task, participants earned \$1 because they did not reach the performance threshold. Table 3.D.3 reports the marginal effects from the Poisson regression for participants in the Effort treatment in which the dependent variable is the number of boxes collected in the BRET. Independent variables are similar to the ones used in the previous tables. The findings in Table 3.D.3 suggest that low performers did not take risk differently than high performers in this treatment. While we interpreted risk taking by the \$2 group in the Effort treatment in terms of entitlement effect (compared to players in the Windfall treatment), this finding suggests that even low performers considered their endowment as hard-earned money to be used conservatively (relatively to the other treatments).

**Compensating for the forgone unethical gain:** The difference in risk taking in the BRET of honest individuals in the Lying and the low performers in the Effort treatment might be explained by the difference in their status quo in part 1. In the CEO task, honest individuals were presented

---

<sup>14</sup>The difference between the marginal effects of Lying and Windfall treatments is significant (Wald test,  $p < 0.001$ ), while that between Lying and Higher Moral Cost treatment is not (Wald test,  $p = 0.155$ ).

Table 3.D.2: Windfall Treatment: \$1 vs. \$2 group

(1)		
Dep. Var.		
Nb. of boxes	ME	St.Err.
\$2 group	-4.229**	(1.953)
Mean SOEP	5.824***	(1.004)
Male	4.217**	(2.002)
Age	-2.447**	(1.024)
Above high school	-1.088	(2.124)
Avg. weekly expenditures	0.003	(0.01)
<i>N</i>	364	
Log pseudo-likelihood	-3474.17	
AIC	6962.34	
BIC	6989.62	

*Notes:* The table reports the marginal effects from Poisson regression.

Delta method standard errors are in parentheses. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 3.D.3: Effort Treatment: \$1 vs. \$2 group

(1)		
Dep. Var.		
Nb. of boxes	ME	St.Err.
\$2 group	2.875	(2.248)
Mean SOEP	5.008***	(1.318)
Male	0.243	(2.278)
Age	-0.864	(1.199)
Above high school	-4.245*	(2.507)
Avg. weekly expenditure	0.015*	(0.009)
<i>N</i>	298	
Log pseudo-likelihood	-3072.18	
AIC	6158.36	
BIC	6184.24	

*Notes:* The table reports marginal effects from Poisson regressions.

Delta method standard errors are in parentheses. \*  $p < 0.10$ , \*\*\*  $p < 0.01$

with the lying option to earn \$2, while low performers had to encode words to reach the threshold to earn \$2. To some extent, honest players *gave up* \$1, while lower performers did not give up anything. This could imply a stronger motive for honest individuals to compensate for the gain foregone in the CEO task by taking more risk in the BRET. Table 3.D.4 reports the marginal effects from Poisson regressions, comparing honest and dishonest individuals in the Lying and Higher Moral Cost treatments. There is no significant difference in the risk taking of honest and dishonest individuals in any of these treatments. The fact that honest individuals took as much risk as the dishonest group who earned more may imply that the former compensated for having foregone the financial gain of being dishonest.

Table 3.D.4: Lying and Higher Moral Cost treatments: \$1 *vs.* \$2 group

Dep. Var.	(1)		(2)	
	Lying		Higher MC	
Nb. of boxes	ME	St.Err.	ME	St.Err.
\$2 group	-1.199	(2.194)	0.751	(2.118)
Mean SOEP	2.681**	(1.071)	4.612***	(1.201)
Male	4.174**	(2.112)	-0.383	(2.241)
Age	-3.008**	(1.271)	-0.992	(1.161)
Above high school	-1.485	(2.485)	0.897	(2.300)
Avg. weekly expenditure	0.005	(0.009)	0.003	(0.009)
<i>N</i>	371		354	
Log pseudo-likelihood	-4055.89		-3803.75	
AIC	8125.78		7621.51	
BIC	8153.19		7648.60	

*Notes:* The table reports marginal effects from Poisson regressions. Delta method standard errors are in parentheses. Mean of responses in the SOEP questionnaire and age are reported in z-scores. \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

# General Conclusion

This thesis contributes to the advances in behavioral economics of dishonesty. It aims at understanding the extent to which a *moral cost*, which arises from individual intrinsic motivations, influences the decision-making process in a moral domain using the methodology of experimental economics. In particular, this thesis focused on lying behavior – When do people lie, and why? How can we use the moral cost to design soft interventions to discourage the likelihood people engage in misbehavior? Under what circumstances do these interventions become ineffective? Chapter 1 and 2 answer these questions in two economically relevant contexts – the former in a financial market setting, where lying is strategic and generates negative externality, and the latter in a group context, where individuals can communicate with each other.

Chapter 1 examines the impact of honesty oaths in the financial markets using a laboratory experiment. Specifically, it investigates whether and how an oath affects advisors' lying behavior, and consequently measures its impact on the decisions of the investors. The novel contribution of Chapter 1 is by varying whether the oath is common knowledge with the investors or not. We find that introducing a common-knowledge oath leads to a drastic reduction in all types of lies and consequently promotes the investment rates. We have identified that the increased investment is driven, not by investors' anticipation of advisors' honesty due to an awareness of the oath, but by the experience of honesty created by a common-knowledge oath. As such, it means that, for the investors, oath has a marginal informational value because they can learn from their past experience whether advisors can be trusted, and respond to it by increasing their investment. Yet, this does not imply that the common-knowledge element is not *necessary* because advisors engage in more lies when the oath is not common knowledge with investors, leading to a lower level of honesty that does not maintain trust. This implies that, from advisors' side, they need to be informed that investors are aware about it in order to bring about a *sufficient* level of honesty and thereby trust in their behaviors. This has important implications for how an oath or a compulsory code of conduct should be implemented in the financial markets: it must be accompanied with a consistent and

well-planned public-relations strategy to ensure awareness both inside and outside the organizations, which eventually brings forth an experience of honesty for the clients and maintain their trust in the sector.

Continuing in a similar vein, Chapter 2 examines the impact of an honesty oath on lying when individuals are in a group setting. Even though evidence in the literature have shown that an oath promotes truth-telling and discourages lying in an individual setting, none of the existing studies have explored if the impact of an honesty oath extends to a group context, where individuals can communicate and may need to make ethical decisions together. This chapter is the first to answer this question using a laboratory experiment. In addition, it explores the phenomenon by varying the incentive structures for the group members (*i.e.*, whether or not group members earn a common payoff). The main contribution of this chapter is by showing that oath intervention may interact with group incentive – when the incentive scheme creates pressure on individuals to conform to the group, it can *crowd out* the desirable impact of the oath. The main finding of Chapter 2 is in line with this – oath reduces lying in group but only when the incentive structure does not create group pressure (*i.e.*, no payoff commonality). The impact of the oath disappears when group members are required to coordinate on an action to earn a common payoff. Analysis of individual lying behavior after group interaction provides an evidence consistent with the group pressure explanation – those who were in the group that faced the payoff commonality rule were less likely to lie later on in an individual setting. The findings of Chapter 2 have implications pertaining to how an oath intervention should be implemented in an organizational context: it must be designed in close connection with the incentive structure.

Chapter 1 and 2 provide evidence that oath can be used as a soft intervention to raise the moral cost and thereby discourages lying behavior in market and group settings, respectively. In addition, they identify the conditions under which the impact of oath may be limited, which provide implications regarding optimal policy designs to reduce fraudulent behaviors. Chapter 3 takes a different perspective by examining whether a moral cost of lying continues to persist *after* individuals have lied and, if so, manifests itself in the way people make economic decisions with the money earned dishonestly. More precisely, Chapter 3 investigates whether and how individuals use the money earned through dishonest means in the context of risk-taking behavior. Do they treat unethically earned money more like a windfall gain from luck or as a result of costly effort? By exploring this phenomenon, this chapter contributes by providing a novel evidence of a *non-motivated violation* of economic fungibility in relation to the ethical source of the money. That is, a moral cost of lying can generate a sense of entitlement, which then affects the way individuals

use their unethical money. We found that in the context of risk taking, dishonestly earned money is treated like a windfall gain and the effect is prevalent among risk averse people. However, increasing the moral cost *ex-post* eliminates the difference between risk-taking with dishonestly earned money and that from costly effort, which implies that a moral cost can induce a sense of entitlement. The findings of Chapter 3 provide a reflection on risk taking in domains such as financial markets, where exploitation of asymmetric information may be tempting. The experimental findings suggest that increasing the intrinsic cost of lying may contribute to induce a conservative use of money, which can perhaps limit excessive gambling.

### Limitations and Extensions

Naturally, there are limitations to the current works presented in this thesis with possible extensions for future research. The use of laboratory and online experiment should not be regarded as a limitation in itself. Evidence from lab studies come together and can improve our understanding of a big complex world by identifying mechanisms underlying behaviors. In addition, scientific knowledge needs to be further tested through replication studies and in a more challenging arena outside the laboratory. With that said, one should still remain cautious about extrapolating the findings to a more complex environment where there is less control.

In Chapter 1, we used a Stranger Matching Protocol which does not allow for reputation building between advisors and investors. However, one may say that advisor-client relationships in the real world are fiduciary in nature and thus take place for a long term. As shown in [Tergiman and Villeval \(2022\)](#), reputation in itself eradicates detectable lies but does not affect deniable ones. On the other hand, Chapter 1 showed that a common-knowledge oath has a dramatic impact on all types of lies in a setting without any reputation. It would be interesting to explore the interaction between *reputation building* and oath. Can oath further add to the impact of reputation in eradicating lies, especially deniable ones? In addition, a reputation allows the investors to punish advisors who have been caught lying by withholding investment. Will investors punish oath-takers who are caught lying? Another interesting extension is to explore if the impact of oath interventions is compromised by an increase in *economic incentives*. How will individuals under oath solve the moral dilemma? Even though the literature suggest that individual lying behavior does not react much to stakes ([Abeler et al., 2019](#)), in the context of the Announcement Game, advisors can make deniable lies that cannot be immediately detected *ex-post*. This plausibility deniability may provide a moral wiggle room ([Dana et al., 2007](#)), which thereby shields them from moral costs. Thus, advisors under oath may react to an increased incentive by resorting to deniable lies strategy.

In Chapter 2, one limitation concerns the way group members communicate using an anonymous chat. To some extent, the observed effect of oath can be seen as a lower bound due to the anonymous environment because individuals may feel comfortable to convince others to lie or to become complicit with the group. In real setting, however, one can anticipate stronger social image concerns to play a role. For instance, modes of communication (*i.e.*, face-to-face, video/voice calling) can increase the reputation cost of cheating via image-related concerns (Bénabou and Tirole, 2006). Therefore, a natural extension would be to lift the *veil of anonymity* and introduce social image concerns, which can make it more difficult for individuals to engage in unethical collaboration with group members. Another interesting extension is to manipulate the *group identity dimension*. The experimental design of Chapter 2 kept the group identity aspect neutral, but one can imagine its role in a more complex setting. If there is a sense of group identity in the oath (*i.e.*, ‘We’ as opposed to ‘I’), it may strengthen the impact of an oath intervention. This is especially interesting since Bicchieri et al. (2022) has shown that social proximity (*i.e.*, a sense of shared identity with one’s peer) can negate the adverse impact of conformity after exposure to a norm violation of others. However, the sense of group identity may also foster an in-group bias, which may swing in favor of group dishonesty.

Lastly, in Chapter 3, we found that increasing a moral cost *ex-post* to the lying induces a sense of entitlement, causing individuals to treat their dishonestly earned money more like that earned from costly effort. An interesting extension is to explore another type of manipulation to influence the moral cost. We used injunctive norms (*i.e.*, telling liars about what one ought to do) to increase the moral cost of lying *ex-post*, but one could explore the use of other norm message (*i.e.*, telling liars about ‘*what people actually do*’). It could be the case that the empirical norm portrays leniency as many people cheat (at least half in the CEO task), which could possibly reduce the moral cost of lying because liars would feel that it is normal to lie. Another direction would be to examine the phenomenon in a peer effect or in a group setting because unethical money is sometimes earned and shared among individuals in a group. Given that groups are more likely to make a rational decision than individuals (Charness and Sutter, 2012) and that violation of a fungibility principle can be viewed as an *irrational* behavior because a dollar in one account is being treated different from that in another account, groups may be less biased than individuals in treating unethical money.

\*\*\*

In all, this thesis adds to the constantly growing body of research on the behavioral economics of dishonesty – how individuals make moral decisions as *homo moralis* who are morally driven, as opposed to *homo economicus* whose aim is to always maximize own material interest. This behavioral economic perspective resonates with the moral philosophies of Adam Smith and Immanuel Kant, whose thoughts have established morality as a foundation of *human socioeconomic betterment*. As human beings, we do not always strive to do what is materially best for ourselves but we also remain within the boundary prescribed by the social codes and the moral duties. Knowing what is expected of us, we avoid those actions that will cause disappointment in others. Besides our intrinsic motivations, we also pay attention to how others will perceive us in the light of our actions. However, as Smith noted that even though we know these normative standards, our greed and selfishness can become our worst enemy that turn us blind, leading us morally astray. By understanding the influence of moral costs on the way people make decisions in the moral domain, this thesis makes novel contributions, which yield behavioral economic insights in designing policies and institutions to discourage dishonesty under the circumstances where people ignore the moral compass, consciously or not.





# Bibliography

- Abbink, K., B. Irlenbusch, and E. Renner (2000). The moonlighting game: An experimental study on reciprocity and retribution. *Journal of Economic Behavior & Organization* 42(2), 265–277.
- Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. *Journal of Public Economics* 113, 96–104.
- Abeler, J., A. Falk, and F. Kosse (2021). Malleability of Preferences for Honesty.
- Abeler, J. and F. Marklein (2017). Fungibility, Labels, and Consumption. *Journal of the European Economic Association* 15(1), 99–127.
- Abeler, J., D. Nosenzo, and C. Raymond (2019). Preferences for Truth-Telling. *Econometrica* 87(4), 1115–1153.
- Ackert, L. F., N. Charupat, B. K. Church, and R. Deaves (2006). An experimental examination of the house money effect in a multi-period setting. *Experimental Economics* 9(1), 5–16.
- Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84(3), 488–500.
- Allingham, M. G. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics* 1(3-4), 323–338.
- Alm, J., R. Bahl, and M. N. Murray (1993). Audit selection and income tax under-reporting in the tax compliance game. *Journal of Development Economics* 42(1), 1–33.
- Alm, J. and A. Malézieux (2021). 40 years of tax evasion games: a meta-analysis. *Experimental Economics* 24(3), 699–750.
- Alm, J., G. H. McClelland, and W. D. Schulze (1992). Why do people pay taxes? *Journal of Public Economics* 48(1), 21–38.

- 
- Anagol, S., S. Cole, and S. Sarkar (2012). *Understanding the incentives of commissions motivated agents: theory and evidence from the Indian life insurance market*. Harvard Business School Cambridge.
- Andersson, O., H. J. Holm, J.-R. Tyran, and E. Wengström (2014). Deciding for Others Reduces Loss Aversion. *Management Science* 62(1), 29–36.
- Antinyan, A. and Z. Asatryan (2019). Nudging for Tax Compliance: A Meta-Analysis.
- Arkes, H. R., C. A. Joyner, M. V. Pezzo, J. G. Nash, K. Siegel-Jacobs, and E. Stone (1994). The Psychology of Windfall Gains. *Organizational Behavior and Human Decision Processes* 59(3), 331–347.
- Balafoutas, L., S. Czermak, M. Eulerich, and H. Fornwagner (2020). Incentives for Dishonesty: An Experimental Study with Internal Auditors. *Economic Inquiry* 58(2), 764–779.
- Balafoutas, L. and N. Nikiforakis (2012). Norm enforcement in the city: A natural field experiment. *European Economic Review* 56(8), 1773–1785.
- Balafoutas, L., N. Nikiforakis, and B. Rockenbach (2014). Direct and indirect punishment among strangers in the field. *Proceedings of the National Academy of Sciences* 111(45), 15924–15927.
- Barron, K. (2019). Lying to appear honest. Technical report, WZB Discussion Paper.
- Barron, K., A. Kajackaite, and S. Saccardo (2022). Image Concerns and Lying Behavior.
- Battigalli, P. and M. Dufwenberg (2007). Guilt in games. *American Economic Review* 97(2), 170–176.
- Bauer, M., J. Cahlíková, D. Celik-Katreniak, J. Chytilová, L. Cingl, and T. Želinsk (2021). Nastiness in Groups. *CERGE-EI Working Paper Series*.
- Beatty, T. K. M., L. Blow, T. F. Crossley, and C. O’Dea (2014). Cash by any other name? Evidence on labeling from the UK Winter Fuel Payment. *Journal of Public Economics* 118, 86–96.
- Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76(2), 169–217.

- 
- Becker, G. S. and G. J. Stigler (1974). Law Enforcement, Malfeasance, and Compensation of Enforcers. *The Journal of Legal Studies* 3(1), 1–18.
- Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.
- Bicchieri, C., E. Dimant, S. Gächter, and D. Nosenzo (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior* 132, 59–72.
- Blumenthal, M., C. Christian, and J. Slemrod (2001). Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota. *National Tax Journal* 54(1), 125–138.
- Boatright, J. R. (2013). Swearing to be Virtuous: The Prospects of a Banker’s Oath. *Review of Social Economy* 71(2), 140–165.
- Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg Registration and Organization Online Tool. *European Economic Review* 71, 117–120.
- Boonmanunt, S., A. Kajackaite, and S. Meier (2020). Does poverty negate the impact of social norms on cheating? *Games and Economic Behavior* 124, 569–578.
- Bornstein, G. and I. Yaniv (1998). Individual and Group Behavior in the Ultimatum Game: Are Groups More “Rational” Players? *Experimental Economics* 1(1), 101–108.
- Brañas-Garza, P., M. Bucheli, M. P. Espinosa, and T. García-Muñoz (2013). Moral Cleansing and Moral Licences: Experimental Evidence. *Economics & Philosophy* 29(2), 199–212.
- Brouwer, T., F. Galeotti, and M. C. Villeval (2022). Teaching Norms in the Street An Experimental Study.
- Bryan, C. J., G. S. Adams, and B. Monin (2013). When cheating would make you a cheater: implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General* 142(4), 1001.
- Buccioli, A. and M. Piovesan (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology* 32(1), 73–78.
- Bursztyn, L., S. Fiorin, D. Gottlieb, and M. Kanz (2018). Moral Incentives in Credit Card Debt Repayment: Evidence from a Field Experiment. *Journal of Political Economy* 127(4), 1641–1683.

- 
- Bénabou, R. and J. Tirole (2006). Incentives and Prosocial Behavior. *American Economic Review* 96(5), 1652–1678.
- Cagala, T., U. Glogowsky, and J. Rincke (2019). Content Matters: The Effects of Commitment Requests on Truth-Telling. SSRN Scholarly Paper ID 3432445, Social Science Research Network, Rochester, NY.
- Cagala, T., U. Glogowsky, and J. Rincke (2021). Detecting and Preventing Cheating in Exams: Evidence from a Field Experiment. *Journal of Human Resources*, 0620.
- Cai, W., X. Huang, S. Wu, and Y. Kou (2015). Dishonest behavior is not affected by an image of watching eyes. *Evolution and Human Behavior* 36(2), 110–116.
- Cain, D. M., G. Loewenstein, and D. A. Moore (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies* 34(1), 1–25.
- Capraro, V. (2018). Gender Differences in Lying in Sender-Receiver Games: A Meta-Analysis. SSRN Scholarly Paper 2930944, Social Science Research Network, Rochester, NY.
- Castillo, G., L. Choo, and V. Grimm (2022). Are groups always more dishonest than individuals? The case of salient negative externalities. *Journal of Economic Behavior & Organization* 198, 598–611.
- Charness, G. and M. Dufwenberg (2006). Promises and Partnership. *Econometrica* 74(6), 1579–1601.
- Charness, G., D. Masclet, and M. C. Villeval (2014). The dark side of competition for status. *Management Science* 60(1), 38–55.
- Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics* 117(3), 817–869.
- Charness, G. and M. Sutter (2012). Groups Make Better Self-Interested Decisions. *Journal of Economic Perspectives* 26(3), 157–176.
- Charroin, L., B. Fortin, and M. C. Villeval (2021). Homophily, Peer Effects, and Dishonesty.
- Chen, C., J. Chen, and G. He (2017). Immorally obtained principal increases investors’ risk preference. *PLOS ONE* 12(4).
- Cherry, T. L. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology* 22(5), 605–615.

- 
- Cherry, T. L., P. Frykblom, and J. F. Shogren (2002). Hardnose the Dictator. *American Economic Review* 92(4), 1218–1221.
- Cohn, A., E. Fehr, and M. A. Maréchal (2014). Business culture and dishonesty in the banking industry. *Nature* 516(7529), 86–89.
- Cohn, A., T. Gesche, and M. A. Maréchal (2022). Honesty in the Digital Age. *Management Science* 68(2), 827–845.
- Cohn, A. and M. A. Maréchal (2017). Laboratory Measure of Cheating Predicts School Misconduct. *The Economic Journal*.
- Cohn, A., M. A. Maréchal, D. Tannenbaum, and C. L. Zünd (2019). Civic honesty around the globe. *Science*.
- Colzani, P., G. Michailidou, and L. Santos-Pinto (2021). Experimental Evidence on the Transmission of Honesty and Dishonesty: A Stairway to Heaven and a Highway to Hell.
- Conrads, J., B. Irlenbusch, R. M. Rilke, A. Schielke, and G. Walkowitz (2014). Honesty in tournaments. *Economics Letters* 123(1), 90–93.
- Conrads, J., B. Irlenbusch, R. M. Rilke, and G. Walkowitz (2013). Lying and team incentives. *Journal of Economic Psychology* 34, 1–7.
- Corgnet, B., R. Hernán-González, P. Kujal, and D. Porter (2015). The Effect of Earned Versus House Money on Price Bubble Formation in Experimental Asset Markets. *Review of Finance* 19(4), 1455–1488.
- Coricelli, G., M. Joffily, C. Montmarquette, and M. C. Villeval (2010). Cheating, emotions, and rationality: an experiment on tax evasion. *Experimental Economics* 13(2), 226–247.
- Coricelli, G., E. Rusconi, and M. C. Villeval (2014). Tax evasion and emotions: An empirical test of re-integrative shaming theory. *Journal of Economic Psychology* 40, 49–61.
- Cox, J. C. (2002). Trust, reciprocity, and other-regarding preferences: Groups vs. individuals and males vs. females. In *Experimental Business Research*, pp. 331–350. Springer.
- Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431–1451.

- 
- Crosetto, P. and A. Filippin (2013). The “bomb” risk elicitation task. *Journal of Risk and Uncertainty* 47(1), 31–65.
- Dai, Z., F. Galeotti, and M. C. Villeval (2017). Cheating in the Lab Predicts Fraud in the Field: An Experiment in Public Transportation. *Management Science* 64(3), 1081–1100.
- Dana, J., R. A. Weber, and J. X. Kuang (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory* 33(1), 67–80.
- Deck, C., J. Lee, and J. Reyes (2008). Risk attitudes in large stake gambles: evidence from a game show. *Applied Economics* 40(1), 41–52.
- Diekmann, A., W. Przepiorka, and H. Rauhut (2015). Lifting the veil of ignorance: An experiment on the contagiousness of norm violations. *Rationality and Society* 27(3), 309–333.
- Dimant, E. and T. Gesche (2021). Nudging Enforcers: How Norm Perceptions and Motives for Lying Shape Sanctions. SSRN Scholarly Paper 3664995, Social Science Research Network, Rochester, NY.
- Dimant, E. and S. Shalvi (2022). Meta-Nudging Honesty: Past, Present, and Future of the Research Frontier. SSRN Scholarly Paper 4081493, Social Science Research Network, Rochester, NY.
- Dimant, E., G. A. van Kleef, and S. Shalvi (2020). Requiem for a Nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization* 172, 247–266.
- Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association* 9(3), 522–550.
- Dufwenberg, M. and M. A. Dufwenberg (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory* 175, 248–264.
- Dulleck, U. and R. Kerschbamer (2006). On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods. *Journal of Economic Literature* 44(1), 5–42.
- Egan, M., G. Matvos, and A. Seru (2018). The Market for Financial Adviser Misconduct. *Journal of Political Economy* 127(1), 233–295.

- 
- Ellingsen, T. and M. Johannesson (2004). Promises, threats and fairness. *The Economic Journal* 114(495), 397–420.
- Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010). Testing guilt aversion. *Games and Economic Behavior* 68(1), 95–107.
- Ellingsen, T. and R. Östling (2010). When Does Communication Improve Coordination? *American Economic Review* 100(4), 1695–1724.
- Elster, J. (1989). Social Norms and Economic Theory. *Journal of Economic Perspectives* 3(4), 99–117.
- Elster, J. (1996). Rationality and the Emotions. *The Economic Journal* 106(438), 1386–1397.
- Elster, J. (1998). Emotions and Economic Theory. *Journal of Economic Literature* 36(1), 47–74.
- Erat, S. and U. Gneezy (2012). White Lies. *Management Science* 58(4), 723–733.
- Erkal, N., L. Gangadharan, and N. Nikiforakis (2011). Relative Earnings and Giving in a Real-Effort Experiment. *American Economic Review* 101(7), 3330–3348.
- Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics* 133(4), 1645–1692.
- Falk, A., E. Fehr, and U. Fischbacher (2003). On the Nature of Fair Behavior. *Economic Inquiry* 41(1), 20–26.
- Falk, A., E. Fehr, and U. Fischbacher (2008). Testing theories of fairness—Intentions matter. *Games and Economic Behavior* 62(1), 287–303.
- Falk, A. and U. Fischbacher (2006). A theory of reciprocity. *Games and Economic Behavior* 54(2), 293–315.
- Falk, A., T. Neuber, and N. Szech (2020). Diffusion of Being Pivotal and Immoral Outcomes. *The Review of Economic Studies* 87(5), 2205–2229.
- Falk, A. and N. Szech (2013). Morals and Markets. *Science* 340(6133), 707–711.
- Faravelli, M., L. Friesen, and L. Gangadharan (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior & Organization* 110, 160–175.
- Fehr, E. and U. Fischbacher (2004). Social norms and human cooperation. *Trends in Cognitive Sciences* 8(4), 185–190.



- 
- Fehr, E. and S. Gächter (1998). Reciprocity and economics: The economic implications of Homo Reciprocans. *European Economic Review* 42(3), 845–859.
- Fehr, E. and S. Gächter (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Review* 90(4), 980–994.
- Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics* 114(3), 817–868.
- Fehr-Duda, H., A. Bruhin, T. Epper, and R. Schubert (2010). Rationality on the rise: Why relative risk aversion increases with stake size. *Journal of Risk and Uncertainty* 40(2), 147–180.
- Fellner, G., R. Sausgruber, and C. Traxler (2013). Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information. *Journal of the European Economic Association* 11(3), 634–660.
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171–178.
- Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in Disguise—an Experimental Study on Cheating. *Journal of the European Economic Association* 11(3), 525–547.
- Foerster, S., J. T. Linnainmaa, B. T. Melzer, and A. Previtro (2017). Retail Financial Advice: Does One Size Fit All? *The Journal of Finance* 72(4), 1441–1482.
- Fosgaard, T. R., L. G. Hansen, and M. Piovesan (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization* 93, 279–284.
- Frey, B. S., M. Benz, and A. Stutzer (2004). Introducing Procedural Utility: Not Only What, but Also How Matters. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft* 160(3), 377–401.
- Fries, T., U. Gneezy, A. Kajackaite, and D. Parra (2021). Observability and lying. *Journal of Economic Behavior & Organization* 189, 132–149.
- Füllbrunn, S. C. and W. J. Luhan (2017). Decision making for others: The case of loss aversion. *Economics Letters* 161, 154–156.

- 
- Galeotti, F., C. Saucet, and M. C. Villeval (2020). Unethical amnesia responds more to instrumental than to hedonic motives. *Proceedings of the National Academy of Sciences* 117(41), 25423–25428.
- Gangadharan, L., N. Nikiforakis, and M. C. Villeval (2017). Normative conflict and the limits of self-governance in heterogeneous populations. *European Economic Review* 100, 143–156.
- Garbarino, E., R. Slonim, and M. C. Villeval (2019). Loss aversion and lying behavior. *Journal of Economic Behavior & Organization* 158, 379–393.
- Gerlach, P., K. Teodorescu, and R. Hertwig (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin* 145(1), 1.
- Gibson, R., C. Tanner, and A. F. Wagner (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review* 103(1), 532–48.
- Gino, F., S. Ayal, and D. Ariely (2009). Contagion and Differentiation in Unethical Behavior: The Effect of One Bad Apple on the Barrel. *Psychological Science* 20(3), 393–398.
- Gino, F., S. Ayal, and D. Ariely (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior & Organization* 93, 285–292.
- Gino, F. and C. Mogilner (2014). Time, money, and morality. *Psychological Science* 25(2), 414–421.
- Gintis, H. (2016). Homo Ludens: Social rationality and political behavior. *Journal of Economic Behavior & Organization* 126, 95–109.
- Gintis, H., S. Bowles, R. Boyd, and E. Fehr (2005). *Moral sentiments and material interests: Origins, evidence, and consequences*.
- Gintis, H. and D. Helbing (2015). Homo socialis: An analytical core for sociological theory. *Review of Behavioral Economics* 2(1-2), 1–59.
- Gioia, F. (2017). Peer effects on risk behaviour: the importance of group identity. *Experimental Economics* 20(1), 100–129.
- Glätzle-Rützler, D. and P. Lergertporer (2015). Lying and age: An experimental study. *Journal of Economic Psychology* 46, 12–25.

- 
- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review* 95(1), 384–394.
- Gneezy, U., A. Imas, and K. Madarász (2014). Conscience Accounting: Emotion Dynamics and Social Behavior. *Management Science* 60(11), 2645–2658.
- Gneezy, U. and A. Kajackaite (2020). Externalities, stakes, and lying. *Journal of Economic Behavior & Organization* 178, 629–643.
- Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review* 108(2), 419–453.
- Grolleau, G., M. G. Kocher, and A. Sutan (2016). Cheating and Loss Aversion: Do People Cheat More to Avoid a Loss? *Management Science* 62(12), 3428–3438.
- Gross, J., M. Leib, T. Offerman, and S. Shalvi (2018). Ethical Free Riding: When Honest People Find Dishonest Partners. *Psychological Science* 29(12), 1956–1968.
- Guiso, L., P. Sapienza, and L. Zingales (2008). Trusting the Stock Market. *The Journal of Finance* 63(6), 2557–2600.
- Guiso, L., P. Sapienza, and L. Zingales (2018). Time varying risk aversion. *Journal of Financial Economics* 128(3), 403–421.
- Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics* 148, 14–31.
- Hastings, J. and J. M. Shapiro (2018). How Are SNAP Benefits Spent? Evidence from a Retail Panel. *American Economic Review* 108(12), 3493–3540.
- Hastings, J. S. and J. M. Shapiro (2013). Fungibility and Consumer Choice: Evidence from Commodity Price Shocks. *The Quarterly Journal of Economics* 128(4), 1449–1498.
- Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 931–959.
- Heinicke, F., S. Rosenkranz, and U. Weitzel (2019). The effect of pledges on the distribution of lying behavior: An online experiment. *Journal of Economic Psychology* 73, 136–151.
- Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. OUP Oxford.

- 
- Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review* 91(2), 73–78.
- Henrich, J., R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich, C. Lesorogol, F. Marlowe, D. Tracer, and J. Ziker (2006). Costly Punishment Across Human Societies. *Science* 312(5781), 1767–1770.
- Hoffman, E., K. McCabe, K. Shachat, and V. Smith (1994). Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior* 7(3), 346–380.
- Holt, C. A. and S. K. Laury (2002). Risk Aversion and Incentive Effects. *American Economic Review* 92(5), 1644–1655.
- Holt, C. A. and S. K. Laury (2005). Risk Aversion and Incentive Effects: New Data without Order Effects. *American Economic Review* 95(3), 902–912.
- Huber, C. and J. Huber (2020). Bad bankers no more? Truth-telling and (dis)honesty in the finance industry. *Journal of Economic Behavior & Organization* 180, 472–493.
- Hvide, H. K., J. Lee, and T. Odean (2019). Easy Money, Cheap Talk, or Spuds: Inducing Risk Aversion in Economics Experiments. SSRN Scholarly Paper ID 3433380, Social Science Research Network, Rochester, NY.
- Imas, A. (2016). The Realization Effect: Risk-Taking after Realized versus Paper Losses. *American Economic Review* 106(8), 2086–2109.
- Imas, A., G. Loewenstein, and C. K. Morewedge (2020). Mental Money Laundering: A Motivated Violation of Fungibility. *Journal of the European Economic Association* (jvaa059).
- Innes, R. and A. Mitra (2013). Is dishonesty contagious? *Economic Inquiry* 51(1), 722–734.
- Irlenbusch, B. and M. C. Villeval (2015). Behavioral ethics: how psychology influenced economics and how economics might inform psychology? *Current Opinion in Psychology* 6, 87–92.
- Jaber-López, T., A. García-Gallego, P. Perakakis, and N. Georgantzis (2014). Physiological and behavioral patterns of corruption. *Frontiers in Behavioral Neuroscience* 8.

- 
- Jacquemet, N., A. James, S. Luchini, J. Murphy, and J. Shogren (2021). Do truth-telling oaths improve honesty in crowd-working? *PLoS ONE* 16.
- Jacquemet, N., S. Luchini, A. Malézieux, and J. F. Shogren (2020). Who'll stop lying under oath? Empirical evidence from tax evasion games. *European Economic Review* 124, 103369.
- Jacquemet, N., S. Luchini, J. Rosaz, and J. F. Shogren (2019). Truth Telling Under Oath. *Management Science* 65(1), 426–438.
- Jacquemet, N., S. Luchini, J. Rosaz, and J. F. Shogren (2021). Can we commit future managers to honesty? *Frontiers in Psychology*, 2785.
- Jacquemet, N., S. Luchini, J. F. Shogren, and A. Zylbersztejn (2018). Coordination with communication under oath. *Experimental Economics* 21(3), 627–649.
- Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization* 93, 328–336.
- Johnson, E., S. Meier, and O. Toubia (2015). Money left on the kitchen table: Exploring sluggish mortgage refinancing using administrative data, surveys, and field experiments. *Working paper* 2(2).
- Kajackaite, A. and U. Gneezy (2017). Incentives and cheating. *Games and Economic Behavior* 102, 433–444.
- Kant, I. (2011). *Immanuel Kant: Groundwork of the Metaphysics of Morals: A German–English edition*. The Cambridge Kant German-English Edition. Cambridge: Cambridge University Press.
- Kantšukov, M. and D. Medvedskaja (2013). From Dishonesty to Disaster: The Reasons and Consequences of Rogue Traders' Fraudulent Behavior. In T. Vissak and M. Vadi (Eds.), *(Dis)Honesty in Management*, Volume 10 of *Advanced Series in Management*, pp. 147–165.
- Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies* 76(4), 1359–1395.
- Kartik, N., M. Ottaviani, and F. Squintani (2007). Credulity, lies, and costly talk. *Journal of Economic theory* 134(1), 93–116.
- Kartik, N., O. Tercieux, and R. Holden (2014). Simple mechanisms and preferences for honesty. *Games and Economic Behavior* 83, 284–290.

- 
- Keizer, K., S. Lindenberg, and L. Steg (2008). The Spreading of Disorder. *Science* 322(5908), 1681–1685.
- Kessler, J. B. and S. Leider (2012). Norms and Contracting. *Management Science* 58(1), 62–77.
- Khalmetski, K. and D. Sliwka (2019). Disguising Lies—Image Concerns and Partial Lying in Cheating Games. *American Economic Journal: Microeconomics* 11(4), 79–110.
- Kiesler, C. A. and J. Sakumura (1966). A test of a model for commitment. *Journal of Personality and Social Psychology* 3(3), 349–353.
- Kimbrough, E. O. and A. Vostroknutov (2016). Norms Make Preferences Social. *Journal of the European Economic Association* 14(3), 608–638.
- Kling, L., C. König-Kersting, and S. T. Trautmann (2019). Investing for Others: Principals ‘vs. Agents’ Preferences. Technical report, Mimeo.
- Kocher, M. G., S. Schudy, and L. Spantig (2018). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science* 64(9), 3995–4008.
- Kocher, M. G. and M. Sutter (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *The Economic Journal* 115(500), 200–223.
- Koessler, A.-K., B. Torgler, L. P. Feld, and B. S. Frey (2019). Commitment to pay taxes: Results from field and laboratory experiments. *European Economic Review* 115, 78–98.
- Kooreman, P. (2000). The Labeling Effect of a Child Benefit System. *American Economic Review* 90(3), 571–583.
- Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). The Formation of Prosociality: Causal Evidence on the Role of Social Environment. *Journal of Political Economy* 128(2), 434–467.
- Kristal, A. S., A. V. Whillans, M. H. Bazerman, F. Gino, L. L. Shu, N. Mazar, and D. Ariely (2020). Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences* 117(13), 7103–7107.
- Kroher, M. and T. Wolbring (2015). Social control, social learning, and cheating: Evidence from lab and online experiments on dishonesty. *Social Science Research* 53, 311–324.

- 
- Krupka, E. L., S. Leider, and M. Jiang (2017). A Meeting of the Minds: Informal Agreements and Social Norms. *Management Science* 63(6), 1708–1729.
- Krupka, E. L. and R. A. Weber (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association* 11(3), 495–524.
- Kugler, T., G. Bornstein, M. G. Kocher, and M. Sutter (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic Psychology* 28(6), 646–657.
- Köbis, N. C., B. Verschuere, Y. Bereby-Meyer, D. Rand, and S. Shalvi (2019). Intuitive Honesty Versus Dishonesty: Meta-Analytic Evidence. *Perspectives on Psychological Science* 14(5), 778–796.
- Lefebvre, M., P. Pestieau, A. Riedl, and M. C. Villeval (2015). Tax evasion and social information: an experiment in Belgium, France, and the Netherlands. *International Tax and Public Finance* 22(3), 401–425.
- Lefebvre, M., F. M. Vieider, and M. C. Villeval (2010). Incentive effects on risk attitude in small probability prospects. *Economics Letters* 109(2), 115–120.
- Levitt, S. D. and J. A. List (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *The Journal of Economic Perspectives* 21(2), 153–174.
- Linnainmaa, J. T., B. T. Melzer, and A. Previtro (2021). The Misguided Beliefs of Financial Advisors. *The Journal of Finance* 76(2), 587–621.
- List, J. A. and T. L. Cherry (2008). Examining the role of fairness in high stakes allocation decisions. *Journal of Economic Behavior & Organization* 65(1), 1–8.
- Loonen, T. and M. R. Rutgers (2017). Swearing to be a good banker: Perceptions of the obligatory banker’s oath in the Netherlands. *Journal of Banking Regulation* 18(1), 28–47.
- Maggian, V. and M. C. Villeval (2016). Social preferences and lying aversion in children. *Experimental Economics* 19(3), 663–685.
- Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013). Poverty Impedes Cognitive Function. *Science* 341(6149), 976–980.
- Marlowe, F. W., J. C. Berbesque, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, J. Ensminger, M. Gurven, E. Gwako, J. Henrich, N. Henrich, C. Lesorogol,

- 
- R. McElreath, and D. Tracer (2008). More ‘altruistic’ punishment in larger societies. *Proceedings of the Royal Society B: Biological Sciences* 275(1634), 587–592.
- Masclet, D., C. Noussair, S. Tucker, and M.-C. Villeval (2003). Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. *American Economic Review* 93(1), 366–380.
- Matsushima, H. (2008). Role of honesty in full implementation. *Journal of Economic Theory* 139(1), 353–359.
- Mazar, N., O. Amir, and D. Ariely (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research* 45(6), 633–644.
- Mazar, N. and C.-B. Zhong (2010). Do green products make us better people? *Psychological Science* 21(4), 494–498.
- Miranda, A. and S. Rabe-Hesketh (2006). Maximum Likelihood Estimation of Endogenous Switching and Sample Selection Models for Binary, Ordinal, and Count Variables. *The Stata Journal* 6(3), 285–308.
- Monin, B. and D. T. Miller (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology* 81(1), 33.
- Mullainathan, S., M. Noeth, and A. Schoar (2012). The market for financial advice: An audit study.
- Nguyen, Y. and C. N. Noussair (2014). Risk Aversion and Emotions. *Pacific Economic Review* 19(3), 296–312.
- Nielsen, K. (2019). Dynamic risk preferences under realized and paper outcomes. *Journal of Economic Behavior & Organization* 161, 68–78.
- Nieper, A. S., C. Huber, T. Promann, and C. Litsios (2021). On Social Norms and Observability in (Dis)honest Behavior.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives* 14(3), 137–158.
- Oxoby, R. J. and J. Spraggon (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization* 65(3), 703–713.
- Peer, E. and Y. Feldman (2021). Honesty pledges for the behaviorally-based regulation of dishonesty. *Journal of European Public Policy* 0(0), 1–21.



- 
- Ploner, M. and T. Regner (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization* 93, 374–383.
- Polinsky, A. M. and S. Shavell (1979). The Optimal Tradeoff between the Probability and Magnitude of Fines. *American Economic Review* 69(5), 880–891.
- Pool, V. K., C. Sialm, and I. Stefanescu (2016). It pays to set the menu: Mutual fund investment options in 401 (k) plans. *The Journal of Finance* 71(4), 1779–1812.
- Prima, R. A., S. Feeny, R. Hoffmann, and E. Satriawan (2020). Threats, Pledges, And Asset Misreporting: A Framed Field Experiment In Indonesia.
- Pruckner, G. J. and R. Sausgruber (2013). Honesty on the Streets: A Field Study on Newspaper Purchasing. *Journal of the European Economic Association* 11(3), 661–679.
- Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review* 83(5), 1281–1302.
- Rahwan, Z., O. P. Hauser, E. Kochanowska, and B. Fasolo (2018). High stakes: A little more cheating, a lot less charity. *Journal of Economic Behavior & Organization* 152, 276–295.
- Sapienza, P. and L. Zingales (2012). A Trust Crisis. *International Review of Finance* 12(2), 123–131.
- Schindler, S. and S. Pfattheicher (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology* 69, 172–177.
- Sen, A. (1973). Behaviour and the Concept of Preference. *Economica* 40(159), 241–259.
- Sen, A. (1994). The Formulation of Rational Choice. *American Economic Review* 84(2), 385–390.
- Sen, A. (2004). *Rationality and Freedom*. Harvard University Press.
- Sen, A. K. (1977). Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs* 6(4), 317–344.
- Shah, A. K., S. Mullainathan, and E. Shafir (2012). Some Consequences of Having Too Little. *Science* 338(6107), 682–685.
- Shalvi, S. (2012). Dishonestly increasing the likelihood of winning. *Judgment and Decision Making* 7(3), 292.

- 
- Shalvi, S., F. Gino, R. Barkan, and S. Ayal (2015). Self-Serving Justifications: Doing Wrong and Feeling Moral. *Current Directions in Psychological Science* 24(2), 125–130.
- Shefrin, H. and M. Statman (1985). The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *The Journal of Finance* 40(3), 777–790.
- Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics* 79(3), 455–483.
- Smith, A. (1759). *The Theory of Moral Sentiments and on the Origins of Languages* (Stewart ed.). London: Henry G. Bohn.
- Suhonen, N. and J. Saastamoinen (2017). How Do Prior Gains and Losses Affect Subsequent Risk Taking? New Evidence from Individual-Level Horse Race Bets. *Management Science* 64(6), 2797–2808.
- Sutter, M. (2009). Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams. *The Economic Journal* 119(534), 47–60.
- Tasimi, A. and S. A. Gelman (2017). Dirty Money: The Role of Moral History in Economic Judgments. *Cognitive Science* 41(S3), 523–544.
- Tergiman, C. and M. C. Villeval (2022). The Way People Lie in Markets: Detectable vs. Deniable Lies.
- Thaler, R. (1985). Mental Accounting and Consumer Choice. *Marketing Science* 4(3), 199–214.
- Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making* 12(3), 183–206.
- Thaler, R. H. (2018). From cashews to nudges: The evolution of behavioral economics. *American Economic Review* 108(6), 1265–87.
- Thaler, R. H. and E. J. Johnson (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science* 36(6), 643–660.
- Thielmann, I. and B. E. Hilbig (2019). No gain without pain: The psychological costs of dishonesty. *Journal of Economic Psychology* 71, 126–137.

- 
- Vanberg, C. (2008a). Why Do People Keep Their Promises? An Experimental Test of Two Explanations<sup>1</sup>. *Econometrica* 76(6), 1467–1480.
- Vanberg, V. J. (2008b). On the Economics of Moral Preferences. *The American Journal of Economics and Sociology* 67(4), 605–628.
- Vieider, F. M., C. Villegas-Palacio, P. Martinsson, and M. Mejía (2016). Risk Taking for Oneself and Others: A Structural Model Approach. *Economic Inquiry* 54(2), 879–894.
- Vincent, L. C., K. J. Emich, and J. A. Goncalo (2013). Stretching the moral gray zone: Positive affect, moral disengagement, and dishonesty. *Psychological Science* 24(4), 595–599.
- Weber, M. and C. F. Camerer (1998). The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organization* 33(2), 167–184.
- Weisel, O. and S. Shalvi (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences* 112(34), 10651–10656.
- Weitzel, U. and M. Kirchler (2021). The Banker’s Oath And Financial Advice. SSRN Scholarly Paper ID 3834397, Social Science Research Network, Rochester, NY.
- Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* 69(3), 309–312.
- Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes* 115(2), 157–168.

---

# Essays on the Behavioral Economics of Dishonesty

Sorravich Kingsuwankul

## Abstract

The aim of this thesis is to investigate the influence of a moral cost, arising from individual intrinsic motivations, on the decision to lie and whether such a cost continues to influence the decision-making process after one has lied. It provides novel evidence on the extent to which the moral cost can discourage lying, the circumstances under which its impact may be limited, and how it can affect decision-making in a risk domain.

Chapter 1 investigates the impact of an honesty oath against lying in a financial market setting using a laboratory experiment. The results show that a common knowledge about the oath (that is, the awareness by all financial advisors and investors) is a key element for its effectiveness. It drastically reduces the prevalence of lies, consequently boosting the level of investment. When oath is not known to the investors, advisors lie more, creating an experience that does not maintain trust. As for the investors, their behaviors are determined by past experience, and not by being aware of the oath.

Chapter 2 examines whether an honesty oath reduces lying in groups where individuals can communicate with each other. Results from a laboratory experiment reveal that the impact of an oath in a group setting depends on the incentive structure. Oath reduces lying only when the incentive structure does not create pressure on individuals to conform to the group. The data show that under the incentive where group members need to coordinate on an action, there is group pressure that crowds out the impact of oath.

Chapter 3 investigates whether individuals take risk with unethically earned money more as a windfall gain or as the result of costly effort. Results from an online experiment reveal that individuals treat dishonestly earned money more like a windfall gain from luck by taking high risk with it. The effect is especially prevalent among risk averse liars. However, increasing the moral cost of lying ex-post eliminates the difference between risk taking with dishonest money and with money earned from costly effort, suggesting that moral costs can generate a sense of entitlement.

Altogether, these findings show that individuals account for the moral cost of lying in their decision-making process, yielding implications for policy and institutional designs to discourage dishonesty using moral costs.

**Keywords:** Lying, Moral Cost, Oath, Risk-Taking, Experiment



---

# Essais en Économie Comportementale de la Malhonnêteté

Sorravich Kingsuwankul

## Résumé

L'objectif de cette thèse est d'étudier l'influence du coût moral, découlant de motivations individuelles intrinsèques, sur la décision de mentir et si un tel coût continue à influencer le processus de prise de décision après avoir menti. Elle fournit de nouvelles preuves sur la mesure dans laquelle le coût moral peut limiter le mensonge, les circonstances dans lesquelles son influence peut être limitée, et comment il peut affecter la prise de décision dans un domaine de risque.

Le chapitre 1 étudie l'impact d'un serment d'honnêteté contre le mensonge dans un contexte de marché financier en utilisant une expérience de laboratoire. Les résultats montrent qu'une connaissance commune du serment (c'est-à-dire une sensibilisation de tous, conseillers financiers comme investisseurs) est un élément clé de son efficacité. Elle réduit considérablement la prévalence des mensonges, ce qui augmente le niveau d'investissement. Lorsque le serment n'est pas connu des investisseurs, les conseillers mentent davantage, créant ainsi une expérience ne permettant pas de maintenir la confiance. En effet : le comportement des investisseurs est déterminé par leur expérience passée, et non par la connaissance du serment.

Le chapitre 2 examine si un serment d'honnêteté réduit les mensonges dans les groupes où les individus peuvent communiquer entre eux. Les résultats d'une expérience en laboratoire révèlent que l'impact d'un serment sur le mensonge en groupe dépend de la structure d'incitation. Le serment réduit le mensonge uniquement lorsque la structure d'incitation ne crée pas de pression sur les individus pour qu'ils se conforment au groupe. Les données montrent que lorsque les membres d'un groupe doivent se coordonner pour obtenir un gain commun, cette incitation crée une pression de groupe qui annule l'impact du serment.

Le chapitre 3 cherche à savoir si les individus prennent des risques avec de l'argent gagné de manière non éthique comme s'il s'agissait d'un gain inattendu ou plutôt comme le résultat d'un effort coûteux. Les résultats d'une expérience en ligne révèlent que les individus traitent l'argent gagné de manière malhonnête davantage comme un gain inattendu dû à la chance en prenant des risques élevés. Cet effet est particulièrement marqué chez les menteurs ayant une aversion pour le risque. Cependant, augmenter le coût moral du mensonge ex-post élimine la différence entre la prise de risque avec de l'argent malhonnête et avec de l'argent gagné au prix d'un effort coûteux, ce qui suggère que les coûts moraux peuvent générer un sentiment de légitimité.

Dans l'ensemble, cette thèse montre que les individus tiennent compte du coût moral du mensonge dans leur processus de prise de décision, ce qui a des répercussions sur les politiques et les institutions visant à décourager la malhonnêteté en utilisant un coût moral.

**Mots Clés:** Mensonge, Coût moral, Serment, Prise de risque, Expérience