Université Paris Nanterre

200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

Membre de l'université Paris Lumières

École doctorale 396 :
Économie, Organisation et Société
Laboratoire :
EconomiX (UMR 7235)

Claire MOLLIER

Comportements et Perceptions à l'épreuve de la comparaison sociale

Expériences en laboratoire et sur le terrain

Thèse présentée et soutenue publiquement le 26 novembre 2024
en vue de l'obtention du doctorat de Sciences économiques de l'Université Paris Nanterre
sous la direction de Monsieur Tarek JABER-LOPEZ
(CSIC, Instituto de Políticas y Bienes Públicos)

Jury :

Examinatrice :	Madame Marie-Pierre DARGNIES	Maître de conférences, Université Paris Dauphine
Présidente :	Madame Johanna ETNER	Professeure, Université Paris Nanterre
Directeur de thèse :	Monsieur Tarek JABER-LOPEZ	Chargé de recherche, CSIC et Instituto de Políticas y Bienes Públicos
Examinateur :	Monsieur Nicolas JACQUEMET	Professeur, Université Paris 1 Panthéon-Sorbonne
Rapporteur :	Monsieur Mathieu LEFEBVRE	Professeur, Université de Strasbourg
Rapportrice :	Madame Sabrina TEYSSIER	Directrice de recherche, INRAE Grenoble

À toutes les femmes de ma vie,

Remerciements

Mon directeur, lors de la finalisation de cette thèse, m'a dit : "tu ne vas pas passer un bel été, tu vas devoir bosser". Celui-ci conclut pourtant trois des plus belles années de ma vie, qui l'ont surtout été puisque bien entourée. L'écriture de ces dernières lignes (pourtant premières dans la thèse) est dédiée à toutes les personnes qui ont contribué à mon épanouissement pendant ces années de doctorat.

Je souhaite en premier lieu remercier la personne sans qui je ne me serais jamais lancée dans cette aventure : Monsieur Tarek Jaber-Lopez. Tarek, tu as su être un directeur exigeant mais toujours bienveillant. Tu m'as laissé ma liberté et mon autonomie sans pour autant me délaisser. Merci de m'avoir transmis le goût de l'économie expérimentale en master. Merci de m'avoir fait confiance en acceptant d'être mon directeur en doctorat. Merci de toujours garder un œil sur moi. On lit souvent dans les remerciements que le choix du directeur est important. Pour moi il était plus que primordial car je n'aurais jamais pu (voire même voulu) faire de thèse, cette thèse, sans ta direction. Gracias por hablarme en mi lengua materna durante los 3 últimos años. (No?) gracias por los interminables podcasts que han alimentado mi lista de tareas pendientes. Gracias por asegurarte de que no me perdiera del todo.

Mes remerciements s'adressent ensuite à mes membres de jury qui me font l'honneur de lire et discuter ma thèse. Mathieu Lefebvre et Sabrina Teyssier, je vous remercie d'avoir accepté de rapporter mes tout premiers travaux de chercheuse. Je suis impatiente de pouvoir améliorer cette thèse grâce à vos retours. Je suis également reconnaissante de pouvoir bénéficier de l'expertise de Marie-Pierre Dargnies et Nicolas Jacquemet en tant qu'examinateur.rice.s. Marie-Pierre, tu es l'une des premières chercheuses que j'ai eu

l'occasion de rencontrer en doctorat. Merci pour le temps que tu m'as accordé, notamment lors mon CSI, ainsi que pour les séminaires et symposium auxquels j'ai pu assister grâce à toi. Merci également à Nicolas qui a pris le temps de relire mes travaux, parfois très peu aboutis, et dont les conseils m'ont été précieux au cours de cette thèse. Enfin, j'exprime ma gratitude envers la dernière membre de mon jury, que j'ai également eu la chance d'avoir en tant que professeure en master. Johanna, merci d'avoir accepté de faire partie de mon jury de thèse. Merci également pour la confiance accordée en master et pour les discussions rassurantes lors de mon doctorat.

Je souhaite remercier le second membre de mon CSI (qui a été à l'occasion discutant de mon séminaire doctorante), dont l'aide précieuse m'a notamment permis d'appréhender les RDD. Benjamin Monnery, merci pour ta disponibilité et tes conseils avisés. Je tiens à remercier l'ensemble de l'équipe pédagogique du master ECA/APE et plus particulièrement Olivier Renault et Ludovic Julien pour leur bienveillance et leur magnanimité à mon égard. D'autres professeures, bien que non expérimentalistes, ont résolument contribué à mon bien-être de doctorante. Valérie Mignon et Cécile Couharde, merci de m'avoir fait profiter de votre sympathie et générosité. Magali Dumontet et Noémi Berlin, merci pour vos enseignements, votre disponibilité et votre jovialité. Merci également aux deux ingénieurs de recherche qui m'ont permis de mener à bien mes projets d'expériences : Fabien Cerrotti et Pierre Vigneron.

J'exprime ma gratitude envers la direction du laboratoire : Nadine Levratto et Saïd Souam. Nadine, merci de nous avoir accordé votre confiance pour l'organisation de séminaires avec Emmanuelle. Saïd, merci de votre confiance et de votre aide qui m'ont permis d'appréhender un peu plus sereinement mes premiers cours en tant qu'enseignante. Mes remerciements s'adressent également à l'ensemble de l'équipe administrative d'EconomiX et plus particulièrement à Abdou, Frédéric, Nasam et Catherine dont la bonne humeur et le sourire ont grandement contribué à adoucir mes lundis matins. Merci aux appariteurs du bâtiment G de toujours secourir mon paquet de chips aux distributeurs.

Au cours de mon doctorat, j'ai eu la chance pouvoir bénéficier de co-écritures enrichissantes

auprès de chercheuses inspirantes. Aurora García-Gallego, muchas gracias por tu acogida en Castellon y por tu generosidad que solo hacen que queramos vivir cerca de Benicassim. Sarah Zaccagni, grazie mille per la tua gentilezza e i tuoi consigli, che saranno preziosi nella mia vita di ricercatrice. Beatrice Braut, grazie per aver condiviso con me la tua conoscenza dell'economia del vino e per la tua simpatia. En dehors de mes co-écritures, j'ai pu enrichir mes travaux au travers de conversations passionnantes avec Emmanuel Peterlé, Helena Fornwagner, Loukas Balafoutas et Paolo Crosetto.

L'école a toujours été un lieu de rencontres et de joie pour moi, c'est donc naturellement que je me tourne vers mes camarades et ami.e.s de doctorat. Merci à ceux et à celles que j'ai eu l'occasion de rencontrer lors des PJEM : Andrea, Éric, Fatou, Laurence et Damien. I would like to thank Mehdi Louafi for being a fantastic human being with a great sense of humour that has helped through all the process of this journey.[1] Je tiens à remercier mon premier bureau à EconomiX (en tant que stagiaire) : le bureau 309. Merci pour votre gentillesse et bon courage pour votre thèse Basile, Florian, Lucas, Maryam et Nolwenn. Merci à mon mentor de m'avoir montré la voie vers la soutenance et de m'avoir stimulé intellectuellement sur les questions de dette publique : Docteur Baptiste Bridonneau. Toujours au troisième étage, je remercie Aliénor pour sa joie de vivre en conférence.

Je monte ensuite de trois étages (avec l'ascenseur de gauche) pour m'adresser à mes voisin.e.s dans le bureau 601. Auparavant, je vous prie d'accepter mes excuses pour tous les rires entendus et provoqués (j'y viendrais) par mes camarades de 602. Du bureau mitoyen au mien je souhaite remercier Docteur Florian Morvillier qui m'a aidé à appréhender les cours de Python que nous avons donnés ensemble lors de ma première année. Un grand merci à Docteur Chouaib Jouf envers qui les mots me manquent pour témoigner de ma reconnaissance. Chouaib, merci de ne m'avoir jamais fait passer au tableau en tant que professeur. Merci de m'avoir pourtant aidé à préparer ma première présentation de thèse en tant que collègue. Et enfin merci pour toutes nos discussions et ton aide précieuse en tant qu'ami. Merci Docteure Capucine Nobletz pour tes conseils autour de verres à mon

[1] Mehdi Besti, merci d'avoir écrit tes propres remerciements.

arrivée en thèse. Plus fraichement diplomé, je souhaite remercier Docteur Lorenzo Garlanda-Longueville pour son amour vache et pour n'avoir jamais manqué d'avoir un mot gentil à mon égard chaque matin (sauf lorsque ma tenue lui déplaisait). Merci à (futur Docteur) Pablo Aguilar Perez pour les découvertes culinaires. Promis, tu ne seras plus jamais "GPI". Merci à Tanguy pour les lasagnes. Merci à Pauline Bucciarelli pour la bonne humeur et les pas de danse. Merci à Hugo Oriola pour ses conseils avisés concernant la conception de mes premiers cours magistraux en tant qu'ATER. Encore bravo pour la validation de ton stage ! J'adresse mes sincères félicitations à la nouvelle génération, et leur souhaite par la même occasion bien du courage : Clara, Eliot et Émilie.

Petit retour en arrière pour adresser un mot à mes anciennes professeures et ami.e.s d'école. Madame Ma Mung et Madame Hauray, merci de m'avoir insufflé le goût de l'économie au lycée. C'est en grande partie grâce à vous que j'ai eu envie d'explorer cette discipline. Je souhaite remercier ma plus vieille amie, depuis le collège : Thelma. Merci à celles qui m'ont suivi au lycée (et surtout après) : Chiara, Inès, Isa, Jeanne B., Jeanne C., Jeanne D. et Zoé. Une mention d'honneur à Inès et Zoé qui ont assisté à mes toutes premières présentations blanches, ainsi qu'à mon bonheur de recevoir la validation des chapitres par le susmentionné Tarek Jaber-Lopez. Vous avez été ma bouée. Merci également particulièrement à Jeanne pour ses mots encourageants. Passons à la fac avec Chloé, dont la douceur et l'attention m'ont permis de garder le moral de la licence au doctorat. Je me dirige ainsi vers le master, dans lequel je souhaite remercier mes anciens camarades : Augustin, Erwan et Thou. Une pensée émouvante envers ceux et celles qui m'ont accompagné jusqu'en doctorat : Arielle, Daniela, Olivier et Pascal. Daniela, merci de m'avoir fait rire à en faire trembler les murs du bâtiment G. Olivier, merci de m'avoir permis d'aller à mon premier cours le cœur plus léger à l'idée qu'un ami était prêt à me sauver[2].

Attaquons-nous à l'une des parties les plus redoutées, annoncant la fin de cette aventure et le début des au revoir. Pour me faciliter la tâche, je débuterais ces remerciements adressés au 602 par celles et celui qui sont déjà parti.e.s. Docteur Morel, merci de m'avoir

[2]J'admire ton inventivité magré l'absence de pipette dans les laboratoires d'économie.

confectionné une petite antisèche pour me présenter à mon premier cours. Merci pour ton affection. Docteure Djénèba, merci de nous montrer que la vie ne s'arrête pas après un doctorat (la preuve : nous pouvons en faire un deuxième !). Merci pour ton sourire. Docteure Fanny, merci de m'avoir emmenée danser et bruncher[3]. Tu ne te rends pas compte à quel point ta soutenance et tes paroles rassurantes m'ont été utiles dans les moments de stress. Docteure María, merci d'avoir été aussi accueillante et d'être aussi dévouée envers l'économie expérimentale. Ta passion et bonne humeur ont fait du bureau 602 un endroit chaleureux. Un autre expérimentaliste, bientôt Docteur, que je souhaite remercier est Moustapha. Ma vie, merci pour ton rire et ta compagnie en conférences. Passons ensuite à ceux qui restent, ceux qui resteront "petits" tout en devenant de grands chercheurs. Albin, merci pour la pertinence de tes relectures et la créativité de tes histoires. Merci d'être têtu, merci de m'avoir fait rire à en pleurer. Aswin, merci de m'avoir relu et de t'être essayer à concevoir avec moi un modèle théorique. Thomas, merci d'avoir agi comme un éclair au cours de ma troisième année. Merci pour les ballons de mon anniversaire, les verres interminables et les rires innombrables. Ton affection est inestimable. Docteure Mathilde, merci pour tes petites attentions, tes mots d'amour qui ont parsemé mon chemin vers le dépôt. Merci de m'envoyer des cartes postales pour un peu atténuer le *blues* de la rentrée. Merci de répondre "Yees ! Je peux faire ça d'ici 10-20min" lorsque je te demande de me relire. Merci de t'émerveiller de chaque avancée mineure dans ma thèse, merci d'être aussi émue que moi quand je m'apprête à déposer. Merci d'être toujours là même si tu n'es plus en face de moi dans le bureau. Merci d'être l'une de mes plus belles rencontres, de ton amitié malgré des débuts mitigés. Concluant la partie sur les membres du 602, j'adresse mes dernières lignes à ma camarade de thèse et voisine de bureau. Emma.nuelle, je n'aurais voulu passer ces 3 années de doctorat avec personne d'autre que toi. Merci de te synchroniser sur la même musique que moi pour que nos après-midis soient plus équilibrés. Merci pour la douceur (quoique souvent chargés) de nos *rendez-vous* en terrasse. Je te suis reconnaissante de m'accorder ton amitié et la chance d'être aux premières loges de ta réussite. Sache que, I had the time of my life fighting dragons with you.

[3]Fanny est une ex bobo parisienne repentie qui vit maintenant à Nancy.

Pour finir, je souhaite m'adresser à ma famille, amies de la famille et aux personnes avec qui je vis. Véronique, merci de m'avoir incité à ranger ma chambre. C'est bien évidemment cela qui m'a permis d'avoir les idées au clair et d'avancer sur mes travaux de recherche. Papa, merci de ne jamais douter de mes réussites et de croire en mes objectifs. Marraine, merci de souvent y croire plus que moi et de toutes les affections dont tu m'as submergée. Inès B., merci pour ta candeur et ta douceur qui font de toi la meilleure collocataire ! Pas si vite, cette catégorie est également disputée par un doctorant du 601 (personne n'est parfait) qui a amplement été présent, autant dans ma vie professionnelle que personnelle. Benjamin, Petit Benji (la liste est longue), merci d'avoir relu la quasi totalité de mes travaux et merci pour tes encouragements. Merci de vouloir descendre à CDG, de lancer le prochain épisode de Fleabag ou encore de me forcer à regarder le Diable s'habille en Prada quand j'ai une petite peine. Merci d'être toujours de bonne humeur, ta présence a été plus qu'appaisante en cette fin de thèse. Merci à ma petite soeur Marie pour sa pointe de folie, sa persévérance et pour m'avoir montré qu'aucun projet n'est assez fou. Merci d'être toujours là même quand tu es loin. Maman, merci de représenter le meilleur role modele qu'une jeune fille, adulte et thésarde puisse rêver d'avoir. Tu es ma source principale de motivation et tu sais toujours avoir les mots justes. Merci aussi à Dodi. Je vous aime plus que tout. Pour finir, j'ai une pensée très émue pour Robin, à qui je ne cesserais jamais de penser.

À toutes les personnes qui ont contribué de près ou de loin, physiquement ou mentalement, je vous remercie du fond du coeur. Entreprendre une thèse met à rude épreuve nos émotions et son accomplissement n'aurait pas eu la même saveur sans vos encouragements et votre accompagnement. Je tiens à adresser ces dernières lignes à des figures qui m'ont soutenu à travers leur musicalité : Billie Eilish, Juliette Armanet ou encore Taylor Swift. Merci à Gérard Blanc d'avoir été dans mes oreilles pendant l'écriture de ces remerciements. Il ne me reste plus qu'à commencer une autre histoire...

Résumé général

Cette thèse s'inscrit dans un contexte de prise de décision individuelle, influencée par la comparaison sociale. L'ensemble des chapitres se construisent autour de l'impact des pairs dans les stratégies en termes de compétition, redistribution ou encore d'utilisation d'application de suivi de pas.

Le premier chapitre mobilise une expérience en laboratoire pour explorer la réaction des hommes et femmes après une défaite. Les sujets performent une tâche mathématique pendant quatre étapes (sous forme de tournoi et de paiement à la pièce) et doivent choisir si il.elle.s souhaitent répéter la compétition. Une attention particulière est portée sur le rôle du genre de l'opposant.e dans cette décision ainsi que l'écart de points entre les adversaires. Dans notre environnement compétitif, nous ne faisons pas état de différences de genre : les hommes et les femmes semblent prêt.e.s à répéter la compétition de manière équivalente et indépendamment du genre de leur adversaire. Ce résultat est à nuancer cependant avec deux facteurs significatifs : le résultat de la précédente compétition (victoire ou défaite) et les attentes des participant.e.s, notamment si il.elle.s sont déçu.e.s de leurs précédents résultats. Par ailleurs, nous montrons que les hommes sont significativement plus enclins à surpasser une femme après une défaite, comparé à un homologue masculin.

Le second chapitre s'interroge sur la perception des inégalités en présence de comparaisons intergroupes. Plus spécifiquement, dans quelle mesure les individus ont tendance à minimiser les inégalités vécues quand ils observent des disparités plus prononcées dans un autre groupe. Dans cette expérience de laboratoire, les sujets d'un même groupe reçoivent différentes allocations et l'étude se concentre sur le comportement redistributif du plus riche.

Les traitements permettent de divulguer : aucune information, l'allocation du plus riche, du plus pauvre, le niveau d'inégalités ou le choix de ne pas redistribuer dans l'autre groupe. Mes résultats démontrent que la révélation d'un niveau d'inégalités plus important dans un autre groupe atténue significativement la volonté de réduire les inégalités à l'intérieur de son propre groupe.

Le troisième chapitre bénéficie de la collaboration avec la start-up de suivi de pas *WeWard* pour l'implémentation d'un Essai Randomisé Contrôlé (ERC) auprès de plus de 20 000 utilisateur.rice.s. Ce chapitre teste l'efficacité de différents types et durées de rappels sur l'engagement envers l'application et les habitudes de marche. Nous envoyons 3 nouveaux types de messages, soulignant chacun un aspect comportemental précis : les coûts irrécupérables, la comparaison avec ses propres performances ou avec celles de pairs. Les résultats montrent que recevoir des messages portant sur les performances des pairs augmente l'utilisation de l'application. En revanche, aucun traitement n'affecte le nombre de pas marchés.

Enfin, nous analysons les comportements de marche et stratégies d'utilisation de l'application de près de 14 000 utilisateur.rice.s *WeWard* dans un quatrième chapitre. Nos données révèlent que, malgré le fait qu'elles déclarent marcher plus dans des études déclaratives, les femmes marchent quotidiennement et hebdomadairement moins que les hommes. Cependant, les utilisatrices semblent ouvrir plus régulièrement l'application. Ce constat suggère que les femmes sont moins engagées à pratiquer une activité physique mais restent dévouées à enregistrer leurs nombres de pas. Cette tendance pourrait même expliquer les différences de genre enregistrées : les femmes enregistrent mécaniquement plus de faibles nombres de pas comparé aux hommes. Ce chapitre souligne l'importance de prendre en compte non seulement le niveau d'activité physique mais aussi les mécanismes par lesquels elle est enregistrée, en particulier dans le contexte des interventions numériques en matière de santé.

Mots clés : économie expérimentale ; inégalités ; genre ; comparaison

Abstract

This thesis assesses the influence of social comparison in individual decision-making. The chapters respectively discuss the impact of peers on strategies in terms of competition, redistribution, and step-tracking app usage.

The first chapter employs a laboratory experiment to explore the reactions of men and women after a defeat. Subjects perform a math task over four stages (under a tournament and piece-rate payment scheme) and decide if they want to repeat the competition. Particular attention is given to the role of the opponent's gender and the point gap between the competitors. In this competitive environment, we do not observe gender differences: men and women appear equally willing to repeat the competition, regardless of their opponent's gender. Nevertheless, this pattern is nuanced by two significant factors: firstly, the outcomes of participants in the preceding competition (victory vs. defeat) and secondly, participants' expectations, particularly if they are disappointed with their previous results. Furthermore, we show that men try significantly more to outperform a woman after a previous defeat, compared to a male opponent.

The second chapter questions the perception of inequality in the presence of intergroup comparisons. Specifically, to what extent do individuals tend to minimize experienced inequality when they observe more pronounced disparities in another group? To investigate this, a laboratory experiment is conducted where participants are matched into groups of two players. Subjects receive different allocations within their group and different types of information regarding the other group: no information, the wealthiest's allocation, the poorest's allocation, both allocations, and the actions to not reduce inequalities within the

group. Data shows that disclosing both the other group's members allocations, and hence their (higher) level of inequality, negatively impacts the willingness to move towards a more equal situation. Revealing the allocation of the wealthiest, or the other group's decision not to redistribute, enhances the perceived fairness of the allocation received.

The third chapter benefits from a collaboration with a start-up of step-tracking (*We Ward*) for the implementation of a Randomized Controlled Trial (RCT) with over 20,000 users. This chapter aims to test the effectiveness of different types and durations of reminders on app engagement and walking habits. We send three new types of messages, each framing a specific behavioral aspect: sunk costs, comparison with one's own performance, or comparison with peers' performance. The results show that receiving messages about peers' performance increases app usage. However, no treatment affects the number of steps.

Finally, we analyze the walking behaviors and app usage strategies of nearly 14,000 *We-Ward* users in the fourth chapter. Our results reveal that, although they report walking more in previous studies, women walk significantly fewer steps than men, both on a weekly and daily basis. However, women are more likely to open the application regularly compared to men. These findings suggest that while women may engage less in physical activity, their commitment to recording steps through the app remains strong. It may also explain why women walk less on average: they record more low amount of steps compared to men. This underscores the importance of considering not just the amount of physical activity but also the mechanisms through which it is recorded, particularly in the context of digital health interventions.

Keywords: experimental economics; inequality; gender; comparison

Table of Contents

Remerciements **i**

Résumé général **vii**

Abstract **ix**

Introduction générale **1**

1 Gender of the Opponent and Reaction to Competition Outcomes **21**

1.1 Introduction 22

1.2 Experimental Design 27

1.2.1 Task 27

1.2.2 Game Description - Part I 28

1.2.3 Risk Preferences and Gender Stereotypes - Part II 32

1.2.4 Procedure 33

1.3 Hypotheses 33

1.4 Results 36

1.4.1 First entry into competition: Stage 3 38

1.4.2 Second entry into competition: Stage 4a 41

1.4.3 Other analysis: Performance 46

1.5 Conclusions 49

1.A Perception of Names 52

1.B Performance across Stages 52

1.C Stage 4b: WTP to Avoid Competition 53
1.D Additional Analyses 56
1.E Robustness Checks 58
1.F Instructions (translated from French) 62

2 Perceptions of inequality: When the grass is less green on the other side 81
2.1 Introduction 82
2.2 Literature 84
2.3 Experimental Design 88
2.3.1 The task 88
2.3.2 Treatments: Disclosure of others' inequality 89
2.3.3 Questionnaire 90
2.3.4 Procedure 91
2.4 Hypotheses 91
2.5 Results 93
2.5.1 Willingness to move towards equality 95
2.5.2 Fairness perceptions 99
2.6 Discussion 101
2.7 Conclusions 102
2.A Descriptive statistics 105
2.B Additionnal tables 107
2.B.1 Willingness to move towards equality 107
2.B.2 Fairness perceptions 109
2.C Additional analysis: Participants' answers 111
2.D Instructions (translated from Spanish) 115

3 The effects of reminders on engagement and walking: Evidence from a large scale experiment 123
3.1 Introduction 124

3.2 Experimental Design . . . 127
3.2.1 The step-tracking app . . . 127
3.2.2 Procedure and Sample . . . 129
3.2.3 Treatments . . . 131
3.3 Hypotheses . . . 133
3.4 Empirical strategy . . . 135
3.5 Results . . . 136
3.5.1 Engagement with the app . . . 137
3.5.2 Walking habits . . . 146
3.6 Conclusions . . . 155
3.A Additional Figures . . . 157
3.B Additional Tables . . . 160

4 Walking for Rewards: Gender Differences in Response to App-Based Incentives 171

4.1 Introduction . . . 172
4.2 The app features . . . 175
4.3 Results . . . 175
4.3.1 Descriptive statistics . . . 175
4.3.2 Investigation of threshold's effect . . . 176
4.4 Conclusions . . . 180
4.A Background . . . 182
4.B Walking behaviors . . . 184
4.B.1 An overview of gender differences in walking habits . . . 184
4.B.2 Recorded measures . . . 185
4.B.3 Self-reported measures . . . 188
4.C App behavior . . . 190
4.C.1 Additionnal figures . . . 190
4.C.2 Additionnal tables . . . 194

Conclusion générale **197**

Bibliography **205**

Liste des figures **235**

Liste des tableaux **238**

Introduction générale

"*There exists, in the human organism, a drive to evaluate his opinions and his abilities. [...] A person's cognition (his opinions and beliefs) about the situation in which exists and his appraisals of what he is capable of doing (his evaluation of his abilities) will together have bearing on his behavior. [...]* ***people evaluate their opinions and abilities by comparison respectively with the opinions and abilities of others.***" (Festinger, 1954)

Cadre général de la thèse

La théorie de la comparaison sociale, initiée par Festinger (1954) dans le domaine de la psychologie, demeure pertinente dans l'environnement économique comme le relève la thèse d'Easterlin (1974). La croissance économique ne s'accompagne pas nécessairement d'une hausse de bien-être si les individus s'enrichissent à un rythme identique. Paradoxalement, lorsque des disparités persistent, tous les individus d'une société sont susceptibles d'en subir les coûts en termes de bien-être (Layte & Whelan, 2014). La comparaison sociale, sous toutes ses formes, affecte donc notre perception des richesses, de nos compétences et ainsi de nos comportements associés. L'étude de ses impacts apparaît d'autant plus pertinente dans un monde connecté, où l'émergence des réseaux sociaux ou autres médias sociaux favorise les possibilités de comparaisons entre individus, parfois non semblables. La "*surveillance des pairs*" constitue l'une des motivations les plus courantes d'utilisation des réseaux sociaux (Joinson, 2008). Pourtant, cette comparaison peut s'avérer biaisée, dans la mesure où elle repose sur des informations que les individus ont volontairement choisi de révéler.

Par conséquent, les utilisateur.rice.s ont tendance à idéaliser la vie d'autrui, en particulier lorsqu'il.elle.s ne connaissent pas personnellement leurs "pairs" (Chou & Edge, 2012).

Selon Wood (1996), la comparaison sociale se définit comme "*le processus de réflexion sur des informations concernant une ou plusieurs autres personnes en les rapportant à soi-même*" (p.520). Ce processus s'articule comme suit. Les individus acquièrent de l'information sociale, puis l'analyse afin de situer la position de leurs pairs (similaire, supérieure ou inférieure). Il.elle.s peuvent ensuite réagir à l'information sociale sous la forme cognitive, affective ou comportementale. La première catégorie rend par exemple compte des objectifs d'auto-évaluation (Festinger, 1954) alors que la forme affective renvoie plutôt aux émotions telles que la jalousie ou la fierté. L'expression comportementale recense les aspirations d'imitation ou de conformité à un groupe. Ces trois éléments sont étroitement liés dans la mesure où un facteur affectif ou cognitif peut engager une réponse comportementale, représentant l'objet d'étude principal de l'économie comportementale.

Les sciences sociales se distinguent de la psychologie par leur incorporation de la dimension sociale dans leur approche des émotions (Petit, 2022). En effet, l'économie comportementale s'intéresse aux comportements individuels et considère généralement la présence de pairs comme élément déterminant de la prise de décision. C'est le cas de la plupart des jeux utilisés en économie expérimentale : le jeu du bien public (pouvant comporter des passagers clandestins), le dilemne du prisonnier (dont l'issue dépend des stratégies des deux joueur.se.s) ou encore les jeux du dictateur ou de l'ultimatum. Sans nécessairement revendiquer l'étude émotionnelle des individus, elle porte son objet d'analyse sur les comportements individuels. C'est précisement dans cette logique que s'inscrit notre thèse. Les quatre chapitres l'articulant ne prétendent pas interpréter les émotions guidant les comportements qui y sont analysés. En revanche, ils ont vocation à déterminer dans quelle mesure certains comportements sont influencés par la comparaison avec des pairs.

Ainsi, selon la catégorisation établie par l'auteur de la théorie de la comparaison sociale, au cours des quatre chapitres de cette thèse nous nous intéresserons conjointement aux *capac-*

ités (Chapitres 1, 3 et 4) et aux *opinions* d'individus (Chapitre 2). Selon Festinger (1954), les opinions ne peuvent être testables de manière objective alors que les abilités (ou performances) bénéficient d'un critère non-ambigu et ordonnable. Ces dernières sont donc sujettes à la comparaison sociale lorsque les performances des individus peuvent être classées, dans un cadre compétitif (Chapitre 1) ou sportif (Chapitres 3 et 4) par exemple. Dans une autre mesure, la comparaison sociale peut permettre d'excuser certains comportements redistributifs relevant plus de la morale, comme il le sera abordé dans le Chapitre 2.

La suite de l'introduction suivra la structure suivante. Dans une première partie, des éléments de contexte seront apportés pour déterminer quel environnement favorise la comparaison entre les individus et quelles en sont les conséquences. Puis, nous présenterons successivement les différents chapitres composant la thèse.

Éléments de contexte

Faut-il se ressembler ou s'opposer pour se comparer ?

Lorsque l'on s'intéresse aux effets de la comparaison sociale sur les comportements individuels, il convient d'aborder la question de la référence choisie pour cette comparaison. Dans le monde anglophone, l'expression "*Keeping up with the Joneses*" (tirée de la bande dessinée éponyme de Arthur R. "Pop" Momand de 1913)[4] renvoie à la comparaison des individus vis-à-vis de leurs "*voisins*", leur servant ainsi de point de référence en termes de classe sociale ou d'accumulation de biens. En macroéconomie, cette attitude est généralement étudiée à travers des modèles de préférences interdépendantes (e.g., Abel, 1990; Gali, 1994). Le revenu du groupe de référence serait tout aussi déterminant pour le bonheur de l'individu que son propre revenu (Ferrer-i-Carbonell, 2005). L'expression peut également être mobilisée pour

[4]L'histoire de la bande dessinée est centrée autour des tentatives de la famille McGinis pour se conformer au style de vie et aux possessions matérielles de leurs voisins (que le lecteur ou la lectrice ne voit jamais), les "Joneses".

illustrer le sentiment d'injustice ressenti quand un pair (du même statut initial) bénéficie d'une ascension, devenant ainsi *supérieur* à l'individu (Burleigh & Meegan, 2013). Selon Wood & Taylor (1991), les individus n'aspirent pas seulement à *keeping up* mais ils choisissent aussi délibérément leurs *Joneses* afin de satisfaire un objectif de "*valorisation*" ou "*d'amélioration*" de soi. Dans le premier cas, les individus réalisent une comparaison dite "*vers le bas*" (*downward*). La comparaison s'effectue dès lors avec des personnes moins bien positionnées (en termes de revenu par exemple) en vue de se valoriser. À l'inverse, les individus peuvent choisir de se comparer avec plus performants (*upward*) qu'eux pour améliorer leurs propres performances. Dans le modèle de Falk & Knell (2004), la valeur de référence augmente avec les facultés des personnes. Les individus à grande capacité peuvent se permettre de se fixer des objectifs élevés, dans la mesure où ils atteindront nécessairement un niveau relativement haut. Ce résultat implique que les individus ont tendance à se comparer à ceux qui leur sont semblables. C'est également une des hypothèses formulée par Festinger (1954): "*une personne ne tend pas à évaluer ses opinions ou abilités en comparaison avec d'autres personnes trop divergentes d'elle même*" (p.120).

Pourtant, la comparaison *upward* ne nécessite pas de se comparer à semblable et peut compromettre l'image de soi. Franzoi *et al.* (2012) ont exploré les différences de genre relatives aux comparaisons d'ordre physique. Leurs résultats montrent que les femmes sont plus susceptibles de réaliser une comparaison *upward* et autocritique, associée à une estime de soi (corporelle) négative. Ce processus de comparaison implique généralement des standards de beauté (souvent présentés dans les publicités ou les films par exemple) difficiles à atteindre. La théorie de l'objectivisation avance par ailleurs que la culture patriarcale ("*qui réduit le corps de la femme à l'état d'objet sexuel*", p.173) conduit les femmes à intérioriser le point de vue de l'observateur et à percevoir leur corps comme un objet à évaluer (Fredrickson & Roberts, 1997). Ces facteurs culturels sont à même d'expliquer pourquoi la pratique d'une activité physique au sein de la population féminine est plutôt dirigée vers un objectif de perte de poids (e.g., Kilpatrick *et al.*, 2005; Egli *et al.*, 2011). Contrairement à la comparaison sociale, les hommes ont tendance à plutôt s'engager dans une comparaison temporelle future,

à aspirer au futur corps qu'ils auront.

Selon Barnett *et al.* (2010), la comparaison engendre deux types de réactions possibles. Soit les individus choisissent d'agir pour rester en tête du groupe ("*the rat race*") soit ils rejettent cette influence. Ces alternatives font écho aux décisions pouvant être prises, par exemple, dans un cadre compétitif. En économie expérimentale, l'analyse des comportements compétitifs a été permise grâce aux travaux fondateurs de Niederle & Vesterlund (2007). Dans leur expérience de laboratoire, les autrices offrent la possibilité aux sujets d'être rémunérés sous forme de paiement à la pièce (ne dépendant que des performances personnelles) ou sous forme de tournoi. Cette dernière forme livre les individus à l'évaluation de leurs abilités vis-à-vis de leur adversaire. Les résultats du papier, par la suite reproduis à de nombreuses reprises, font état de différences de genre. Les femmes, à l'inverse des sujets masculins, ont tendance à fuir la compétition et à être moins performantes dans les environnements compétitifs (Gneezy *et al.*, 2003). La comparaison sociale y joue un rôle dans la mesure où la confiance des individus en leurs capacités peut découler de stéréotypes de genre, associés à la tâche devant être effectuée. Par exemple, les femmes sont généralement perçues comme moins bonnes en mathématiques que les hommes (Nosek *et al.*, 2002; Guiso *et al.*, 2008; Bordalo *et al.*, 2019). Lors d'un tournoi mixte, si ces stéréotypes sont partagés par l'adversaire masculin, celui-ci aspire alors à une victoire. De manière générale, les individus sont fortement susceptibles de gagner lorsqu'ils concourent avec d'autres individus moins bien classés qu'eux. En cas de défaite, le résultat peut alors ne pas constituer une opportunité d'auto-évaluation, en raison de différences notables avec l'adversaire. Nous nous plaçons alors dans l'alternative du refus de l'influence de la comparaison.

La comparaison sociale peut donc admettre diverses formes en fonction des caractéristiques individuelles ou des objectifs qu'elle vise. À travers l'étude de sa nature, il convient également d'appréhender les effets qu'elle peut produire, notamment en termes de stéréotypes.

La comparaison comme vecteur de stéréotype

Afin d'expliquer les différences de genre en termes de comportements compétitifs, et comprendre si elles sont immuables, les articles distinguent habituellement l'approche par l'inné (*nature*) de celle par l'acquis (*nurture*). La première explication renvoie aux spécificités biologiques et plus particulièrement aux coûts de reproduction (Darwin, 1871; Bateman, 1948; Trivers, 1972; Knight, 2002). Selon la théorie Darwinienne, dans les sociétés barbares, les hommes étaient sujets à la compétition et c'était aux femmes que revenaient le choix d'un géniteur. Cette exposition à la compétition, selon une logique évolutionniste, aurait mené les hommes à avoir une nature plus ambitieuse là où les femmes seraient plus *tendres* et altruistes[5]. Cette approche, dépeignant la nature intrinsèque des Hommes, est discutée par les sciences expérimentales, visant à comprendre si une modification de l'environnement affecte les choix. Ainsi, nous assistons par exemple à l'étude de société matrilinéaire, en comparaison à un système patriarcal, avec Gneezy *et al.* (2009)[6] ou à l'impact de la composition des écoles avec Booth & Nolen (2012). Lorsque les filles étudient dans des écoles non-mixtes, elles ne sont pas plus susceptibles de fuir la compétition que les garçons. Ces résultats renvoient à l'idée que l'identité de genre n'est plus menacée en l'absence du genre opposé.

Les environnements mixtes sont susceptibles de favoriser des comportements conformes aux stéréotypes de genre en raison, notamment, des *backlash effects*. Selon Moss-Racusin & Rudman (2010), ceux-ci se rapportent aux conséquences économiques et sociales d'une attitude non appropriée selon les stéréotypes de genre. En France, 62% des hommes (et 50% des femmes) considèrent que les femmes sont plus *douces* (Haut conseil à l'égalité entre les femmes et les hommes, 2023). Dans ce contexte, il peut apparaître difficile de s'imposer comme une femme compétitive, qui par définition renvoie aux attitudes assertives et non

[5] "*Woman, owing to her maternal instincts, displays these qualities towards her infants in an eminent degree; therefore it is likely that she should often extend them towards her fellow-creatures. Man is the rival of other men; he delights in competition, and this leads to ambition which passes too easily into selfishness.*" (p.326, Darwin, 1871).

[6] La société étudiée dans Gneezy *et al.* (2009) constitue bien un exemple de société matrilinéaire (et non matriarcale) dans la mesure où c'est l'ascendance maternelle qui prime chez les Khasi. En revanche, les femmes n'assument pas pour autant les rôles généralement attribués aux hommes dans les sociétés patriarcales.

coopératives (Kilmann, 1983). Les femmes sont par ailleurs jugées plus arrogantes quand elles communiquent leurs aptitudes (*self-promotion*) ou elles ont moins de chances d'être embauchées (par un responsable masculin) lorsqu'elle négocient leur salaire (Bowles *et al.*, 2007; Rudman *et al.*, 2012). La peur de ces effets *backlash* conduit les femmes à dissimuler (voire minimiser) leurs succès et expliquerait leur moindre réussite en négociation (Amanatullah & Morris, 2010).

Cet effet de découragement se produit également au sein d'autres minorités stigmatisées qui cessent par exemple leur recherche d'emploi par anticipation de la discrimination subie (Heslin *et al.*, 2012). Dans cette situation, la comparaison est d'ordre temporelle. Les individus prennent des décisions en considérant leurs expériences passées. Sur le marché du travail, celle-ci peut également s'apparenter à une comparaison sociale, c'est-à-dire entre différents profils de candidat.e.s. Il s'agit par ailleurs de la façon la plus courante d'identifier des facteurs discriminants, comme le montre la célèbre expérience des CV à l'aveugle initiée par Bertrand & Mullainathan (2004)[7]. Dans l'étude de la discrimination, il convient de présenter les deux modèles qui prévalent, à savoir les modèles de discrimination par goût (*taste-based*) de Becker (1957) et par inférence statistique (*statistical*) de Phelps (1972) et Arrow (1972). La première forme de discrimination renvoie à l'idée que les employeur.se.s jugeraient désagréable de travailler avec certaines minorités et éviteraient donc de les embaucher. La discrimination chez Phelps (1972) et Arrow (1972) admet quant à elle que, pour évaluer les capacités d'un.e candidat.e, les employeur.se.s inférent celles-ci par rapport aux caractéristiques moyennes du groupe auquel l'individu appartient. La réussite de l'individu repose donc à la fois sur une comparaison temporelle (des expériences vécues avec d'autres individus de son groupe) et sociale, puisque appartenant à un certain groupe de référence. Ces mécanismes conduisent ainsi à la définition même de la discrimination, à savoir : "*le traitement injuste et négatif des travailleur.se.s ou des candidat.e.s à un emploi sur la base d'attributs personnels qui ne sont pas pertinents pour juger de la performance au travail*"

[7]L'expérience a par la suite été reproduite à de nombreuses reprises : Edo & Jacquemet (2014) pour le cas français, Blommaert *et al.* (2014) pour les Pays-Bas, McGinnity *et al.* (2009) pour l'Irlande, Wood *et al.* (2009) pour l'Angleterre ou encore Liebkind *et al.* (2016) pour la Finlande.

(p.34, Chung, 2001).

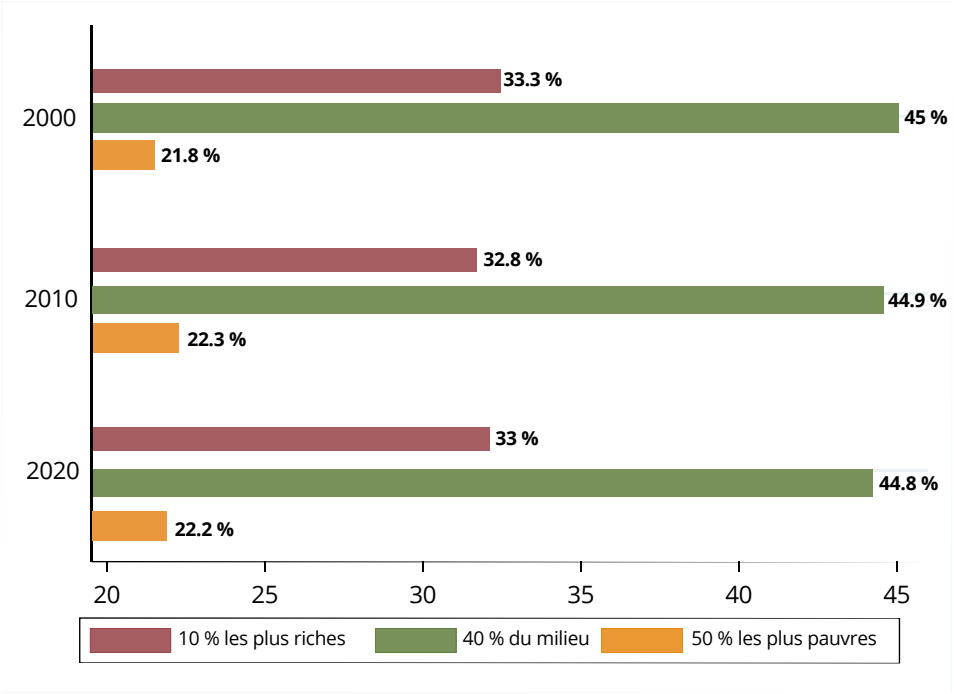
Pour autant, la question du "*traitement injuste*" peut apparaître plus délicate en raison de mécanismes d'auto-réalisation, comme la menace du stéréotype (*stereotype threat*). Cette expression s'apparente à une pression ressentie par les membres d'un groupe stéréotypé lorsque l'individu doit réaliser une tâche dans laquelle son groupe d'appartenance est supposé moins compétent (Aronson *et al.*, 1999). Les individus ont tendance à anticiper qu'ils vont confirmer le préjugé négatif en question, produisant un résultat auto-réalisateur. Leurs performances sont moindres par rapport au groupe bénéficiant du stéréotype positif et, à long-terme, ces stéréotypes (vérifiés) affaiblissent l'estime de soi. Les effets de la menace du stéréotype dépendent essentiellement de la tâche à effectuer et du stéréotype qui est activé (Ambady *et al.*, 2001). Par exemple, le stéréotype négatif le plus commun, associé aux mathématiques, est le fait que les femmes sont moins performantes que les hommes et le versant positif est que les individus d'origine asiatique sont plus doués[8]. Dans leur expérience, Ambady *et al.* (2001) évaluent les effets de l'activation d'un de ces deux stéréotypes sur des enfants. Leurs résultats révèlent que les filles obtiennent de meilleurs résultats lorsque leur identité ethnique (i.e. Asio-Étasunienne) est mise en avant, au moyen d'un dessin. Leurs performances diminuent si c'est leur identité de genre qui est plus saliente. Selon Aronson *et al.* (1999), les membres d'un groupe majoritaire peuvent aussi être victimes de menace du stéréotype mais il.elle.s constituent des cibles indirectes puisqu'il.elle.s représentent justement le "standard". Dans leur expérience, les hommes ont tendance à être moins performants en l'absence d'un stéréotype implicite statuant de leur supériorité en mathématiques face aux femmes. Leurs performances peuvent même être inférieures à celles des femmes lorsqu'il est annoncé qu'ils sont généralement moins performants (Leyens *et al.*, 2000). En présence d'éléments de comparaison, les groupes minoritaires et à la fois majoritaires sont donc influencés par le fait que leur désavantage soit souligné. Dans la thèse, cette situation est étudiée dans le cadre des inégalités au cours du deuxième chapitre.

[8]En 2023, le podium des performances en mathématiques (score PISA) est occupé par Singapour, le Japon et la Corée du Sud.

La comparaison sociale au service des inégalités

Nous venons d'appréhender dans quelle mesure la comparaison pouvait être dommageable pour les individus et comment elle pouvait véhiculer des stéréotypes. Ces derniers, étant à l'origine de certaines discriminations, alimentent les inégalités au sein d'une société. En France, près d'un tiers des revenus totaux sont capturés par les "classes supérieures" (soit les 10 % les plus riches) alors que les 50 % de la population les plus pauvres ne détiennent qu'à peine un quart des revenus (voir Figure 1). Cette situation ne semble pas s'améliorer et les économistes alertent sur le niveau d'inégalités actuel au sein même des pays (Chancel *et al.*, 2022). De cet état de fait, les chercheur.se.s s'intéressent aux conséquences du niveau d'inégalités sur les relations sociales et notamment sur le renforcement de la distance sociale entre les individus (Wilkinson & Pickett, 2017; Willis *et al.*, 2022).

Figure 1: Répartition des revenus en France



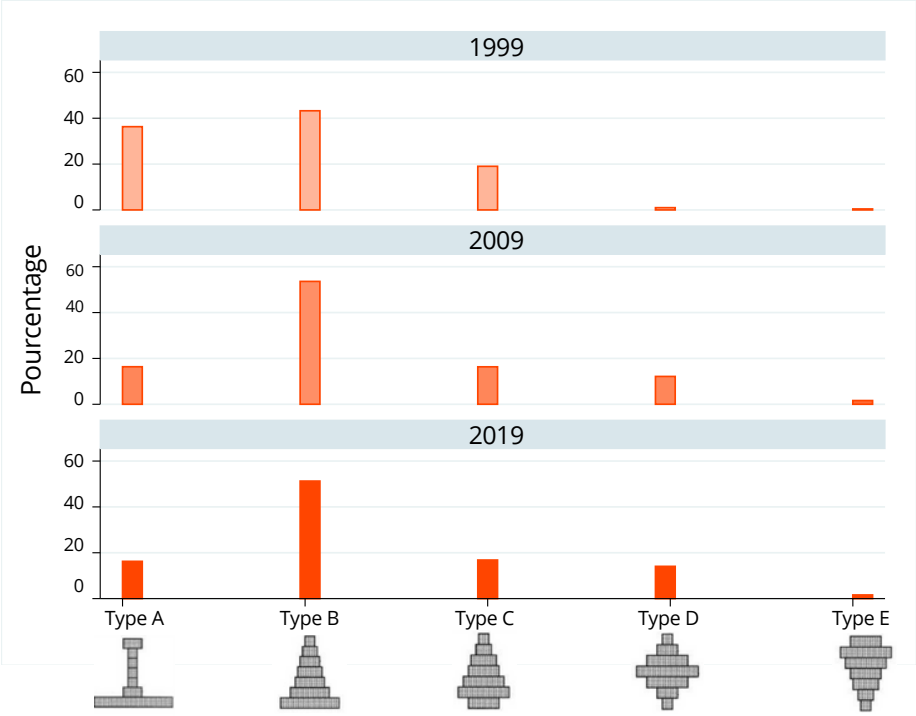
Source : Les données sont issues de l'ouvrage de Piketty (2021) et sont disponibles sur piketty.pse.ens.fr/egalite

Notes : Les revenus comprennent à la fois ceux du travail et ceux liés au capital.

D'après Wilkinson & Pickett (2017), avec un niveau d'inégalités important, les individus sont plus sensibles au statut social et aspirent à se différencier socialement des autres. La distance sociale peut aussi se percevoir en termes de participation à des activités collectives et sociales, à travers la confiance entre les agents ou envers les institutions politiques (e.g., Alesina & La Ferrara, 2000; Uslaner & Brown, 2005). En raison du besoin de différenciation, les inégalités favorisent la compétition et la comparaison sociale entre les individus (Cheung & Lucas, 2016; Sommet *et al.*, 2019). Selon Willis *et al.* (2022), ces résultats suggèrent que les inégalités économiques poussent les individus à se concentrer sur eux-mêmes et à délaisser les autres. Ce faisant, le niveau d'inégalité accroît la distance sociale et réduit donc les liens sociaux ainsi que la coopération. L'article souligne que ces effets découlent du niveau d'inégalités perçu et pas nécessairement vécu, voire même "*si les personnes ne percoivent pas le niveau d'inégalités, elles peuvent percevoir le climat normatif que les inégalités créent, ce qui pourrait produire des effets similaires*" (p.306, Willis *et al.*, 2022). À ce propos, des institutions s'interrogent sur la perception des citoyen.ne.s quant aux inégalités, comme présenté sur la Figure 2. Depuis le début du siècle, la perception d'une moyennisation de la société émerge (type D) mais c'est surtout la vision d'une société sous forme pyramidale (type B) qui semble dominer.

Des inégalités économiques peuvent découler d'autres formes de disparités, comme les inégalités en termes de santé. Selon plusieurs sources, les troubles mentaux se généralisent progressivement (Baxter *et al.*, 2014; Trautmann *et al.*, 2016; Bagein *et al.*, 2022). En France métropolitaine, 1 personne sur 10 présente des symptômes de troubles dépressifs en 2019 et les individus à faible revenu sont deux fois plus touchés (Bagein *et al.*, 2022). C'est dans ce contexte que deux de nos chapitres s'intéressent à une activité physique gratuite (la marche) afin d'encourager les individus à pratiquer une activité physique – bénéfique pour leur santé physique et mentale – sans creuser les inégalités de santé. Le deuxième avantage d'une activité gratuite étant d'ordre plus macroéconomique. Selon Trautmann *et al.* (2016), plus de la moitié de la population à revenu moyen et élevé souffre d'au moins un trouble mental au cours de leur vie. Compte tenu de la progression de l'espérance de vie dans ces mêmes

Figure 2: "À quel type de société correspond la France ?"



Source : Données provenant de l'enquête sur les inégalités sociales de l'International Social Survey Programme (ISSP) pour les années 1999, 2009 et 2019. Les données pour 2019 sont disponibles via : https://data.progedo.fr/studies/doi/10.13144/lil-1522, pour 2009 : https://data.progedo.fr/studies/doi/10.13144/lil-0544 et pour 1999 : https://data.progedo.fr/studies/doi/10.13144/lil-0237.

Notes : Le type A correspond à une société avec une petite élite au sommet, très peu de gens au milieu et la grande masse en bas. Le type B à une forme pyramidale, soit une petite élite en haut, beaucoup de gens au milieu et encore plus en bas. Le type C est similaire à la forme B mais avec moins de personnes en bas. La forme D correspond à une toupie, avec une classe moyenne importante. La forme E représente la situation où beaucoup d'agents sont proches du sommet et très peu sont à la base.

pays, la charge à long terme des troubles mentaux est destinée à croître. La pratique d'une activité régulière permettrait de prévenir ce type de problèmes de santé, sans nécessairement alourdir les dépenses de santé de l'État, mécaniquement plus importantes lorsque le système de sécurité sociale est développé. C'est le cas de la France par exemple, où les dépenses de santé représentaient 9,1 % du Produit Intérieur Brut (PIB) en 2021 (DREES, 2022). Selon l'ISCA & CEBR (2015), si seulement 5 % des personnes inactives pratiquaient une activité

physique, cela permettrait de générer des économies de près de 6,4 millions d'euros pour le système de santé français. Le pays avait par ailleurs récemment proposé une loi sur le remboursement des activités physiques par le système de sécurité sociale afin de réduire les coûts de santé liés à la sédentarité[9].

Dans le contexte de l'activité physique, l'influence des pairs a été démontrée efficace pour encourager les individus à pratiquer une activité régulière. Pratiquer l'activité avec des pairs permettrait de surmonter les obstacles potentiels tels que la fatigue ou de mauvaises conditions météorologiques (Fitzgerald *et al.*, 2012). Ces résultats, combinés à l'utilisation croissante des applications de suivi d'activité physique, conduisent à l'émergence de nouveaux réseaux et applications[10]. Dans ce cadre, la comparaison ne renvoie pas à une démarche de différenciation mais permet plutôt de s'adapter à semblable pour encourager une pratique vertueuse. Cet effet sera considéré dans notre Chapitre 3. Notre travail de thèse, dans son ensemble, a donc pour ambition de penser les effets de la comparaison sociale, autant du point de vue de ses méfaits que de ses bénéfices.

Cette thèse, en mobilisant différents outils de l'économie expérimentale (expérience de laboratoire et de terrain) et économétriques (e.g., analyse en panel, RDD), ambitionne d'analyser les effets de la comparaison sociale sur diverses thématiques. Ce travail commencera par un chapitre portant sur les différences de genre en termes de comportements compétitifs. Puis, nous chercherons à comprendre si la comparaison sociale constitue un obstacle à la poursuite de l'égalité de richesse. Enfin, dans les deux derniers chapitres, il sera question d'activité physique et d'utilisation d'application de suivi d'activité. Nous présenterons les résultats d'une expérience de terrain menée en collaboration avec une start-up de suivi de pas. Le dernier chapitre de cette thèse abordera les différences d'utilisation d'application liée à l'activité physique chez les hommes et les femmes. Les objectifs et contributions sont présentés plus en détails dans la partie qui suit.

[9]Voir la proposition sur le site de l'Assemblée via : *https : //www.assemblee − nationale.fr/dyn/16/textes/l16b0993*$_p$*roposition − loi*

[10]L'application Walk Buddy (https://www.walkbuddy.app/), par exemple, permet de rencontrer des partenaires de marche.

Les objectifs et contributions de la thèse

Chapitre 1 - L'esprit de compétition après une défaite en fonction du genre de l'adversaire

Le Chapitre 1, intitulé "*Gender of the Opponent and Reaction to Competition Outcomes*" propose une expérience de laboratoire, reproduisant une partie du protocole expérimental de Niederle & Vesterlund (2007), afin d'étudier les comportements de compétition des hommes et des femmes. Ce travail présente deux originalités notables. Tout d'abord, il se caractérise par une nouveauté méthodologique dans l'analyse des comportements compétitifs, à savoir l'utilisation d'une méthode stratégique (*strategy-method*). Les sujets effectuent une tâche mathématique sous la forme d'un paiement à la pièce puis d'un tournoi. Ils ont ensuite la possibilité de choisir leur forme de rémunération pour l'étape suivante. Ceux et celles ayant opté pour le tournoi découvrent, à la fin du jeu, si il.elle.s ont gagné le tournoi. L'expérience leur propose alors de rejouer la compétition, en fonction de l'écart de points avec leur adversaire. En d'autres termes, les sujets déterminent à l'avance, pour chaque écart de points potentiels, leur mode de rémunération pour l'étape qui suit. Cette configuration permet d'étendre l'étude des choix compétitifs, au delà d'un choix binaire d'entrée (ou non) en compétition.

L'autre contribution majeure de ce chapitre réside dans son objet d'étude. En effet, notre expérience rend compte des réactions après une défaite ou une victoire, en considérant l'influence du genre de l'opposant.e dans la décision de concourir à nouveau. L'étude de ces deux effets combinés apparait essentielle pour comprendre les mécanismes qui alimentent les inégalités de genre sur le marché du travail. Pour permettre aux participant.e.s d'identifier le genre de leur adversaire sans introduire de biais de demande ou de désirabilité sociale, nous adoptons le protocole utilisé par Datta Gupta *et al.* (2013). Après avoir déclaré leur genre, les participant.e.s sélectionnent un prénom parmi une liste de prénoms masculins ou féminins, correspondant à leur genre déclaré. Avant chaque tournoi, le prénom choisi est

révélé à l'adversaire.

Nos données révèlent des différences de genre dans l'entrée en compétition, ce qui est concordant avec les résultats obtenus dans la littérature. Les sujets, masculins comme féminins, préfèrent légèrement (mais pas significativement) plus concourir face à une femme que face à un homme. Les hommes de notre échantillon ont tendance à plus rejouer la compétition face à un homologue masculin que féminin quand ils ont gagné la précédente compétition. En revanche, après une défaite face à une femme, les hommes optent plus pour le tournoi que lorsqu'ils perdent face à un homme. Ce résultat est seulement significatif pour les hommes sur-confiants : ceux qui s'attendaient à gagner et qui ont perdu. Il semblerait donc que la réaction, chez les hommes, après une défaite face à une femme soit influencée par la déception. Les femmes, quant à elle, concourent plus après avoir gagné et moins lorsque leur sur-confiance leur est révélée. Contrairement aux hommes, ce dernier résultat n'est pas influencé par le genre de leur adversaire. En s'intéressant à l'évolution des performances, nos données rendent compte de différences significatives après une défaite. Les hommes, précédemment battus par une femme et qui ont décidé de rejouer la compétition, semblent améliorer leurs performances.

Chapitre 2 - Les actions en faveur de l'égalité en présence de comparaison sociale

Le deuxième chapitre, "*Perceptions of inequality: When the grass is less green on the other side*" s'intéresse aux déterminants de la volonté de réduire les inégalités, à travers une expérience de laboratoire. Notre protocole expérimental se construit autour d'un jeu du dictateur modifié (Forsythe *et al.*, 1994) dans lequel les individus sont placés dans des groupes de deux. Les deux joueur.se.s reçoivent des allocations initiales différentes et l'individu le plus riche se voit offrir la possibilité de redistribuer les ressources. Il peut donc choisir de maintenir la situation initiale inégalitaire ou au contraire permettre aux deux membres du groupe de percevoir la même allocation. Dans chaque session expérimentale, un autre groupe, plus

inégalitaire, est présent et fait face au même choix. Dans le but d'explorer l'influence de la comparaison sociale sur ce choix, l'expérience comprend plusieurs traitements.

L'hypothèse principale testée, constituant ainsi notre contribution majeure, interroge l'effet d'une comparaison inter-groupe du niveau d'inégalités sur les choix de redistribution. Le premier traitement consiste donc à révéler le niveau d'inégalités vécu dans le groupe le plus inégalitaire au groupe bénéficiant d'un niveau d'inégalités moins important. Pour tester si l'impact engendré par cette information réside dans les croyances que forment les individus sur les actions prises dans le groupe le plus inégalitaire, un second traitement dévoile la décision prise dans le groupe le plus inégalitaire. Enfin, afin de contrôler le point de référence choisi par les sujets dans le processus de comparaison, deux traitements sont ajoutés. Ces derniers divulgent respectivement l'allocation de l'individu le plus riche et celle du plus pauvre du groupe le plus inégalitaire. L'ensemble des résultats obtenus dans les traitements sont analysés en comparaison d'un groupe contrôle, qui ne reçoit aucune information sur l'autre groupe.

Notre résultat principal admet que la comparaison sociale influence les choix de redistribution. En présence de comparaison inter-groupe du niveau d'inégalités, les individus semblent moins enclins à choisir l'option permettant une égalité au sein du groupe. L'absence de significativité dans les autres traitements ne nous permet en revanche pas d'en expliciter les mécanismes. Les données de l'expérience laissent néanmoins suggérer que le processus de comparaison entre groupes d'individus (voire pays) semble être un obstacle dans la poursuite de l'égalité.

Chapitre 3 - L'effet de rappels de la performance des pairs sur la marche et l'utilisation des applications de santé

Pour ce troisième chapitre, "*The effects of reminders on engagement and walking: Evidence from a large scale experiment*", nous bénéficions d'une collaboration avec la start-up de suivi

de pas *WeWard* pour la mise en place d'une expérience de terrain. Cette expérience s'inscrit dans un contexte d'utilisation croissante des applications de suivi d'activité physique et contribue significativement à la littérature en raison de l'ampleur de l'échantillon étudié (près de 20 000 individus). En outre, la collaboration avec l'application nous permet d'implémenter nos traitements auprès d'individus ignorants leur participation à une expérience, limitant ainsi un potentiel biais de désirabilité sociale ou de demande.

Concernant le fonctionnement de l'application, *WeWard* offre à ses utilisateur.rice.s des points (pouvant être convertis en argent) à chaque palier de pas atteint. Afin de maximiser l'engagement et attirer l'attention sur l'activité physique effectuée quotidiennement, les utilisateur.rice.s doivent chaque jour "valider" leurs pas sur l'application. *WeWard* envoie chaque jour à heure fixe des rappels à ses utilisateur.rice.s pour leur rappeler de valider leurs pas. Dans ce travail, nous cherchons à comprendre si des rappels plus ciblés, mettant en avant des aspects comportementaux précis, permettent une meilleure attention envers l'application et, de fait, une hausse de l'activité physique de marche.

L'application, outre la rémunération offerte, permet de suivre ses progrès personnels mais également ceux d'ami.e.s aussi présent.e.s sur *WeWard*. Nos traitements reproduisent ces trois composantes de l'application. Dans les différentes conditions expérimentales, les utilisateur.rice.s ont reçu des messages mettant en avant l'aspect monétaire de l'application, la comparaison avec leurs propres performances passées ou avec celles de pairs. La durée optimale d'implémentation est également étudiée. Certains sujets reçoivent les nouveaux messages pendant 1 semaine et d'autres sujets pendant 3 semaines. Le groupe contrôle reçoit, tout au long de l'expérience, le message générique habituellement envoyé par l'application. Ainsi, nous bénéficions d'une période de pré-traitement de 2 semaines, d'une période de traitement (variant de 1 à 3 semaines) et d'une période de post-traitement (2 à 4 semaines). Cette configuration nous permet de mener une analyse en panel sur un large échantillon.

Les données révèlent que la probabilité d'ouvrir l'application augmente lorsque les messages envoyés mettent en avant la performance des pairs. Ce résultat permet de rendre

compte de l'effet de la comparaison sociale sur l'utilisation de l'application. Il constitue également une marge de manœuvre pour influencer les habitudes de marche et, peut-être, permettre l'implémentation future d'autres traitements. Puisque les utilisateur.rice.s ouvrent plus l'application, il.elle.s prennent plus conscience du nombre de pas effectué. Pourtant, nous ne trouvons pas d'effet significatif de nos traitements sur les habitudes de marche. Les comportements de marche suivent une trajectoire qui dépend du sujet et de caractéristiques externes spécifiques, telles que le genre. Les sujets sont plus sédentaires pendant les week-ends et les vacances, ce qui suggère que la marche est liée aux activités professionnelles et sert davantage de moyen de transport que d'activité de santé récréative.

Chapitre 4 - Différences de stratégies d'utilisation d'application en fonction du genre

Enfin, le dernier chapitre de cette thèse s'intitule "*Walking for Rewards: Gender Differences in Response to App-Based Incentives*". Bien que sa portée soit plus descriptive, il mobilise comme outil économétrique les Regression Discontinuity Design (RDD) pour décrire les habitudes de marche et stratégies d'utilisation d'application de santé des hommes et femmes. La base de données utilisée est celle qui est également exploitée dans le Chapitre 3. De plus, nous mobilisons des données administratives produites par le Centre de Recherche pour l'Étude et l'Observation des Conditions de vie (CREDOC), l'Institut National de la Jeunesse et de l'Éducation Populaire (INJEP) et différents ministères. L'accès à ces différentes sources de données nous permet de rendre compte des divergences qui peuvent exister entre des données déclaratives et celles enregistrées via un système podométrique.

Un des apports majeurs de ce chapitre est la mise en évidence de contradictions entre des données déclaratives et enregistrées. Comme pour le Chapitre 3, l'originalité de la base de données permet de mener à bien un travail économétrique important et vise à enrichir la littérature concernant les analyses de différences de genre dans la pratique d'une activité physique. En plus d'apporter une étude comparative des habitudes de marche, ce travail

rend compte des différences d'utilisation d'application de santé grâce à un RDD permettant d'illustrer les potentielles motivations de validation de pas chez les utilisateurs et utilisatrices.

Nos données *WeWard* révèlent que, malgré le fait qu'elles déclarent marcher plus, les femmes enregistrent un nombre de pas moyen inférieur à celui des hommes. Pourtant, elles semblent plus interagir avec l'application vu qu'elle l'ouvrent plus régulièrement. Nos résultats ne font état d'aucune différence de genre significative en termes de stratégie de validation de pas. Les utilisateurs et utilisatrices, de manière équivalente, sont plus susceptibles de valider leurs pas conditionnellement aux récompenses reçues. Les résultats du chapitre soulignent l'importance de prendre en compte les motifs d'interaction avec les applications de santé, car ils peuvent avoir un impact sur la mise en œuvre de politiques de santé.

Expériences de laboratoire

Chapter 1

Gender of the Opponent and Reaction to Competition Outcomes [1]

"I'm out to prove that a guy 55 years old, with one foot in the grave, can play with the best woman in the world and maybe beat her." – Bobby Riggs (1973)

[1]This chapter is a joint work with Aurora García-Gallego (ICAE and Economics Department, Universidad Complutense de Madrid and LEE and Department of Economics, Universitat Jaume I, Castellón, Spain), Tarek Jaber-Lopez (Institute of Public Goods and Policies (IPP), Spanish National Research Council (CSIC), Madrid, Spain), and Sarah Zaccagni (Department of Economics and Business Economics, Aarhus University, Denmark).

1.1 Introduction

According to the latest report of UN Women (2023), only 28.30% of women held managerial positions worldwide in 2020. Robust evidence from developed countries indicates that this enduring occupational segregation significantly contributes to wage disparities between men and women (Sorensen, 1990; Groshen, 1991; Blau & Kahn, 2017; Cortes & Pan, 2018). Women's access to high positions remains limited, as evidenced by the number of women at the head of a CAC40 company (i.e. only three in 2023).[2] It is well known that occupations are disproportionately distributed by gender, and occupations over-represented by women tend to pay lower wages than occupations over-represented by men (Goldin, 2014). Beyond horizontal segregation, vertical segregation—marked by the concentration of women and men in distinct grades, levels of responsibility, or positions—warrants considerable attention. Among OECD countries, we observe a 19.10% salary differential favoring men in the 9th decile of salary range, compared to a 10.60% difference for those in the bottom salary range (OECD, 2023). These figures give evidence of the existence of the "glass ceiling" effect (Arulampalam *et al.*, 2007). There are several reasons why women do not advance in their career at the same pace as men do. Perceived or experienced discrimination (e.g., Hersch, 2007; Azmat & Petrongolo, 2014) and the trade-off between women's professional careers and personal lives (DiBenedetto & Tittle, 1990) are key factors that contribute to this gender gap. Women's aspirations may be hindered both by self-perception (Beaman *et al.*, 2012), and by other external factors; for instance, their promotion opportunities decrease in contexts where male leaders dominate (De Paola & Scoppa, 2015). The under-representation of women in senior and high-paying positions fuels a vicious circle, discouraging younger women to aspire for similar roles (Ellingrud *et al.*, 2016; Morais Maceira, 2017). People who want to advance in their careers must confront competitive environments and situations, such as admission procedures, job applications, promotion tournaments, and competition between entrepreneurs. Therefore, individual attitude in competitive environments contributes to determine an individual's career trajectory, at least partially (Buser & Yuan, 2019).

[2]CAC40 is the French stock market index that tracks the 40 largest French stocks.

An extensive literature in experimental economics provides evidence of gender differences in competitiveness, and this evidence has become a popular explanation for gender differences in labor market outcomes (e.g Gneezy & Rustichini, 2004; Niederle & Vesterlund, 2007; Balafoutas & Sutter, 2012).[3] Indeed, in the labor market, competitiveness is valuable for accessing promotions or high-paying (and highly competitive) positions. Once documented a relevant gender gap in competitiveness (Niederle & Vesterlund, 2007), several papers have analyzed how different features of the competition (high vs. low stakes, identity of the opponent, repeated vs. one-shot, feedback provision, etc.) or different individual attributes (i.e. family background, age, culture, race, etc.) play a role in it (Dreber *et al.*, 2014; Sutter & Glätzle-Rützler, 2015; Almås *et al.*, 2016; Nicholls, 2022; Hauge *et al.*, 2023).

In this chapter, we focus on the features of the competition, and more specifically, we investigate the effect of two of them combined: the effect of winning or losing in a previous competition and the gender of the opponent. We investigate how these two factors affect subsequent willingness to compete again. We consider this particularly relevant because both features are crucial characteristics of the competitive environment in real life. Indeed, career decisions are made sequentially (i.e. employees apply for promotions repeatedly), and at the end of each competition, candidates are typically informed about the outcome (i.e. whether they obtained the promotion) and are aware of the gender of the competitor. To achieve top positions, what matters is not only subjects' willingness to compete but their willingness to re-compete after losing a single competition. As empirical data show that women are more likely to drop out after experiencing setbacks, especially in male-dominated fields (Price, 2010; Ellis *et al.*, 2016; Buser & Yuan, 2019), it becomes crucial to investigate how men and women react to losing or winning, and whether this is affected by the gender of the competitor, to shed light on mechanisms that rule labor market dynamics.

The milestone experimental study on gender differences in willingness to compete is

[3] A recent strand of literature has put the role of competitiveness under scrutiny. In particular, Gillen *et al.* (2019) show that measurement error in measures of risk attitudes and overconfidence might induce to overestimate the importance of competitiveness in lab experiments. Van Veldhuizen (2022a) re-examine this result using a novel experiment and claims that the entire gender gap is driven by gender differences in risk attitudes and self-confidence.

Niederle & Vesterlund (2007). It documents, in a lab experiment, a large gender gap in tournament entry, driven by men being more overconfident and more willing to perform in a competition. After it, most of the literature on the topic has confirmed that men tend to be more confident and more competitive than women (Kamas & Preston, 2012; Dreber *et al.*, 2014; Ludwig *et al.*, 2017; Alan *et al.*, 2020). The observed trend is generally an over-representation of low-performing men and an under-representation of high-performing women in competition. While in a lab setting, these sub-optimal choices directly impact the payment of subjects, at the macro level, a loss of efficiency is incurred if qualified women are not assigned to high positions (in favor of less able men). The analysis of public interventions, aimed at reducing the gender gap in competitiveness and in the labor market, shows that these differences are not irreversible (Balafoutas & Sutter, 2012; Beaman *et al.*, 2012).

Moving from the individual characteristics to the features of the competition, Buser *et al.* (2023) test and confirm the validity of the result on gender difference in willingness to compete in a quasi-experimental real-world setting with exceptionally high stakes.[4] In this study, the authors analyze also how the willingness to compete is influenced by the gender of the opponent, finding that women tend to avoid competing against men. On the other hand, men seem to anticipate female behavior, and show a larger willingness to compete against women. As in many tasks, there are stereotypes regarding gender differences in performance, it is important to study how the gender of the opponent affects participants' behavior in a competition. In recent experiments in the US and China, He & Noussair (2024) investigate whether men incur a psychological cost when they are outperformed by a female competitor. They find that in China men are willing to pay more to hide that they lost regardless of the gender of the competitor, while there is no gender difference in the US. In China, women are more ashamed to lose to another woman than to a man, while in the US, the opposite occurs. The identity of the opponent is also the focus of Apicella *et al.* (2017), who look at the competition against oneself versus against others and find no evidence of a gender difference in the willingness to self-compete. Both papers suggest the importance of the identity of the

[4] Cai *et al.* (2019) also report that compared to male students, females underperform on the most competitive and high-stakes China's entrance exam.

subject against whom an individual competes.

Another important feature of the competition is whether the decision to compete is or is not one-shot. To better mimic the labor market setting, it is crucial to study not only the willingness to compete in a single competition, but also how this willingness evolves in response to winning and losing in repeated competitions. Empirical evidence has shown that in male-dominated fields, women are more likely than men to drop out after experiencing setbacks (Katz *et al.*, 2006; Goldin, 2015; Fischer, 2017). Some experimental studies test whether winning or losing in a competition affects subjects' behavior in a subsequent competition, and whether men and women react differently in ways that may enhance the gender gap. In a lab experiment, Buser (2016) studies the effect of a competition's outcome on the willingness to seek further challenges and they find gender differences in reaction to winning or losing. Men react to losing by picking a more challenging target while women lower their performance. Gill & Prowse (2014) investigate the effect of winning or losing on later performance and find that women's performance decreases after losing, while for men it does only for high-stakes.

Testing the effect of competition outcomes on behavior in a second competition raises the question of whether feedback induces belief updates. Evidence is inconclusive. Some studies (Ertac & Szentes, 2011; Wozniak *et al.*, 2014) show that precise performance feedback reduces the gender gap in competitiveness. Others (Cason *et al.*, 2010; Möbius *et al.*, 2022) find that the gap persists, with men updating their beliefs less conservatively and being less feedback-averse. Research (Albrecht *et al.*, 2013; Berlin & Dargnies, 2016; Buser *et al.*, 2018; Coutts, 2019) consistently shows that women update beliefs more conservatively and react strongly to personal performance feedback. Baier *et al.* (2024) find that rank feedback encourages more women to compete. Buser & Yuan (2019) reveal that women are less likely than men to compete again after losing. Fang *et al.* (2021) link this gender difference to labor market disparities, noting girls are more likely to drop out after failing in previous competitions, especially in countries with larger labor market gender gaps.

Compared to previous studies, we are the first to combine two features of the competitive environment: the effect of winning or losing in a previous competition and the gender of the opponent. Our primary outcome is the subsequent willingness to compete again. We implement a novel design in which we use a strategy-elicitation method where subjects have to report whether they are willing to compete a second time, based on a sequence of different point gaps between their performance and their opponent's performance. Through this strategic method, we can identify the switching point in subjects' decisions to re-compete. This method provides a continuous and hence more nuanced measure of participants' (dis)taste for competition. At a secondary level, we examine how subjects perform after a winning or losing to investigate potential gender differences in the inclination to outperform an opponent of the opposite (same) gender. In addition to the willingness to re-compete and the performance, we elicit individuals beliefs to examine how they evolve throughout the competition. Indeed, if gender stereotypes are shared by men and women, beliefs about the opponent could play a role in their decision to compete.[5]

Our findings first reveal a gender gap in the taste for competition, with men exhibiting a higher willingness to compete compared to women, in line with previous studies. Both men and women show slightly, albeit statistically non-significant, preference for the tournament payment scheme when competing against female counterparts. After winning against a woman, men are less willing to compete against her again, compared to after winning against a man. After losing against a woman, men are more willing to repeat the competition, compared to after losing against a man. This pattern is only significant for overconfident men.[6] Women compete more after winning and compete less when they are overconfident, regardless of their opponent's gender. This result suggests that the reaction to winning or losing does not necessarily operates through beliefs' updating. Regarding the evolution of performance, we observe significant differences after losing. Men, after losing and re-entering competition against women, substantially improve their performance. Women perform consistently in

[5]In Mago & Razzolini (2019) and Geraldes (2020), women are discouraged to perform against males. In Datta Gupta *et al.* (2013), when men know that they will compete against a woman, they choose more the competition, but this trend is reversed when they think that the woman is abler than them.

[6]Men who expect to win and end up loosing.

both stages, regardless the outcome and the gender of their opponent.

The rest of the chapter is structured as follows. In Section 1.2, we present the experimental design. We report in Section 1.3 our testable hypotheses, and in Section 1.4 the main results. Finally, Section 1.5 concludes the chapter.

1.2 Experimental Design

1.2.1 Task

We conduct an experiment in which participants solve a real task: they are asked to add up sets of five two-digit numbers and fill in the sum in the blank box (Niederle & Vesterlund, 2007). Each subject repeats this task for 3 minutes over four stages (see Subsection 1.2.2 for a detailed description of the stages). Participants can use paper and pens provided, but cannot use their telephone or calculator.[7]

At the beginning of the experiment, subjects fill in a questionnaire with socio-demographic questions (gender, age, and education). Then, they have to choose a name from a list of most given babies' names in 2020 in France (INSEE). Subjects choose from a list of female and male names according to the gender declared in the survey. A survey conducted prior to this experiment allows checking that the first names were unanimously perceived as female or male (see Appendix 1.A).[8] At this point, subjects are randomly matched with an opponent. The matching is made in order to have a sufficient number of participants in each condition: Male vs. Male, Female vs. Female, Male vs. Female and Female vs. Male (henceforth respectively MM, FF, MF and FM). Subjects keep the same opponent all along the experiment. In Stage 2, participants are implicitly informed of the gender of their opponent by the disclosure of

[7]Refer to Subsection 1.2.4 for a description of the procedure and to Appendix 1.F for the complete instructions of the experiment.

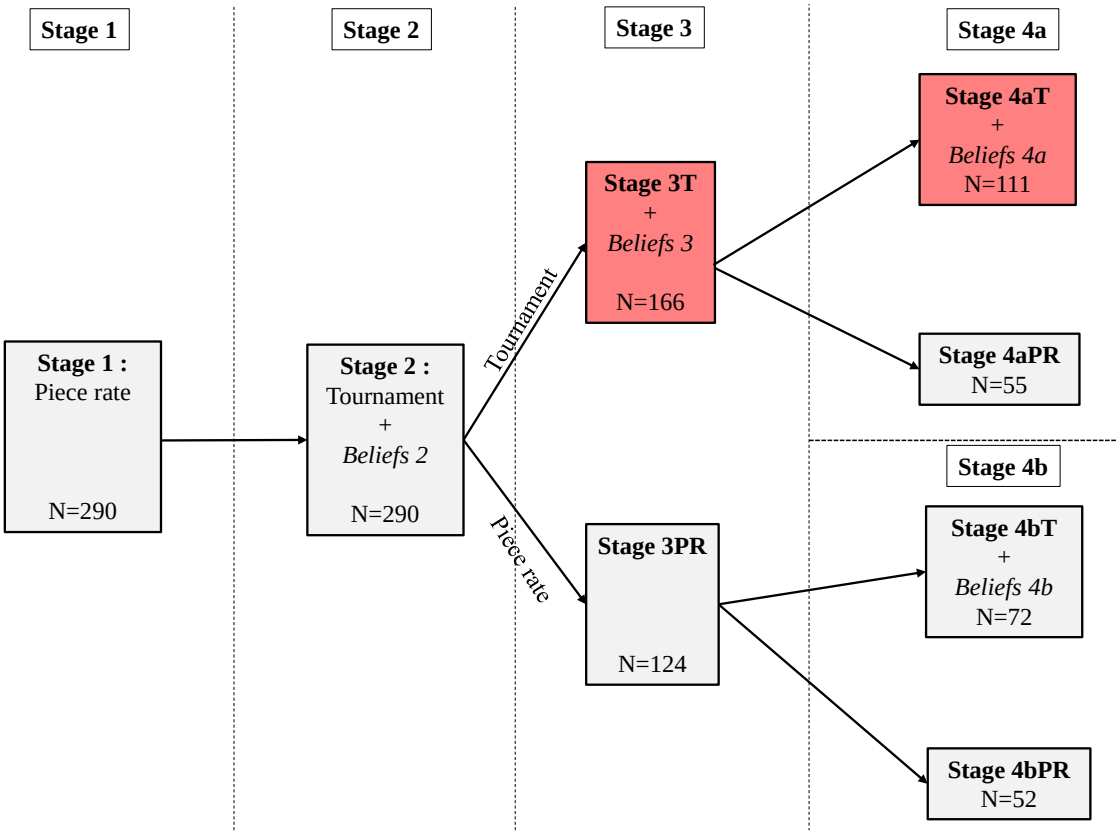
[8]If the participant answered "Others", they are offered a list of male and female names. At the end of the experiment, 75.17% of participants correctly guessed the gender of their opponent.

the name previously chosen by our subjects. Then, they are reminded of their opponent's name at each tournament stages.

1.2.2 Game Description - Part I

In this Section, we describe all stages of the game that participants play in the first part of the experiment (Part I). Figure 1.1 shows the timeline of the game. Refer to Appendix 1.F for the complete instructions of the experiment.

Figure 1.1: Timeline of the experiment (Part I)



Stage 1: Piece rate

In Stage 1, participants perform the task for 3 minutes, under a piece rate payment scheme. They receive 0.50 € for each correct sum. Subjects are informed on their perfor-

mance in this stage at the end of Part I of the experiment. We deprive subjects of feedback in order to reveal preferences towards competition and not rational choices regarding their abilities. It also allows us to elicit participants' confidence in their mathematical skills. Finally, playing without knowledge of past performance reduces the noise during the capture of their willingness to re-compete in the final stage. This absence of feedback prevents subjects from establishing a hypothetical calculation of the point gap with their opponent, and optimizes the implementation of our strategy method.

Stage 2: Tournament

In Stage 2, subjects perform the task for 3 minutes, under a tournament scheme. In this stage we disclose subjects' chosen names at the beginning of the stage. Subjects are randomly paired and compete against each other. In this form of tournament, the best performer receives 1.50 € for each correctly solved problem, while the other receives zero (winner-takes-all tournament). Ties are broken randomly. Subjects are informed on their performance in this stage at the end of the part I of the experiment. At the end of this stage, we elicit subjects' beliefs about their status: subjects guess whether they won or lost. Correct beliefs are rewarded with 1.00 € (henceforth *Beliefs 2*).

Stage 3: Choice of payment-scheme

In Stage 3, subjects decide on their payment scheme, i.e. whether they want to be paid based on a tournament or based on a piece rate. If they opt for the piece rate (henceforth referred to as Stage 3PR), the compensation is determined by the number of correct answers, as in Stage 1. Alternatively, if participants choose the tournament option (henceforth referred to as Stage 3T), their performance is compared to their opponent's performance in Stage 2. If they solve more problems in Stage 3T than their opponent solved in Stage 2, they receive a payment of 1.50 € per correct answer, otherwise they receive 0.00 €. Subjects perform the task for 3 minutes and their remuneration is based on the chosen payment scheme. At the end of this stage, we elicit subjects' beliefs about their status: subjects guess whether they won or lost. Correct beliefs are rewarded with 1.00 € (henceforth *Beliefs 3*). Subjects

are informed on their status in this stage (i.e. whether they won or lost) at the end of this stage. Subjects who choose piece rate are informed about their performance at the end of Part I of the experiment.

After Stage 3, we split our sample based on the chosen payment scheme. Therefore, we can distinguish subjects who show a taste versus distaste for competition. In the subsequent stage, participants who opted for the tournament payment scheme in Stage 3 are assigned to Stage 4a, while those who chose the piece rate in Stage 3 are assigned to Stage 4b.

Stage 4a: Second entry into competition

After receiving feedback on their status (winner/loser) at the end of Stage 3, participants enter Stage 4a. Here, they are presented with the option to decide whether they want to compete again against the same opponent. We use strategy elicitation method by providing them with a sequence of hypothetical points gap between their and their opponent's performance. Subjects observe a scale displaying the performance gap and are given the option to accept the competition (or not) for each gap (see Figure 1.2). Once a decision is made, we disclose the actual point gap between the participant's performance and that of their opponent in Stage 3. If the participant's choice was to decline to compete for the actual point gap, they receive payment according to the piece rate (Stage 4aPR). Alternatively, if the participant's choice was to accept to compete, payment is determined by the tournament payment scheme (Stage 4aT). Subsequently, their performance in Stage 4 is compared to that of their opponent. At the end of this stage, we elicit subjects' beliefs about their status: subjects guess whether they won or lost (henceforth *Beliefs 4a*). Correct beliefs are rewarded with 1.00 €.

Figure 1.2: Example of the performance gap's scale

(a) Winners

Would you "accept or "not accept" to compete again against Louis, according to the following different scenaria from the previous stage?

	I accept	I do not accept
You have solved correctly the same number of sums, so you are tied	o	o
You have solved correctly 1 more sum	o	o
You have solved correctly 2 more sum	o	o
You have solved correctly 3 more sum	o	o
You have solved correctly 4 more sum	o	o
You have solved correctly 5 more sum	o	o
You have solved correctly 6 more sum	o	o
You have solved correctly 7 more sum	o	o
You have solved correctly 8 more sum	o	o
You have solved correctly 9 more sum	o	o
You have solved correctly 10 more sum	o	o

(b) Losers

Would you "accept or "not accept" to compete again against Louis, according to the following different scenaria from the previous stage?

	I accept	I do not accept
You have solved correctly the same number of sums, so you are tied	o	o
Your opponent solved correctly 1 more sum	o	o
Your opponent solved correctly 2 more sum	o	o
Your opponent solved correctly 3 more sum	o	o
Your opponent solved correctly 4 more sum	o	o
Your opponent solved correctly 5 more sum	o	o
Your opponent solved correctly 6 more sum	o	o
Your opponent solved correctly 7 more sum	o	o
Your opponent solved correctly 8 more sum	o	o
Your opponent solved correctly 9 more sum	o	o
Your opponent solved correctly 10 more sum	o	o

Stage 4b: WTP to avoid competition

Participants who choose the piece rate payment scheme in Stage 3 are forced to compete in Stage 4b. Participants can avoid the competition by giving up a certain amount of their total gain. This amount of money corresponds to their willingness to pay (WTP) to avoid the tournament. At the same time, an amount of money is picked randomly by the experimenter; we refer to it as the willingness to pay random (WTP_R). The amount randomly chosen is deducted from the participant's total gains. Based on the subject's choice and the random pick, there are two possibilities:

- If $WTP < WTP_R$ (i.e. the amount chosen by the subject is lower than the amount randomly chosen): subjects have to compete (Stage 4bT). After the tournament, we elicit their beliefs, and correct beliefs are rewarded with 1.00 € (henceforth *Beliefs 4b*).
- If $WTP \geq WTP_R$ (i.e. the amount chosen by the subject is higher than the amount randomly chosen): subjects avoid the tournament (Stage 4bPR).

Stage 4b is designed as an alternative to Stage 4a to further reveal competitive preferences. It highlights the pronounced aversion to competing of some subjects who choose a high level

of WTP and a nuanced taste (or even indifference) for those who choose a low level of WTP. As this stage falls outside the primary scope of our work, results from Stage 4b are not incorporated into the main text but are detailed in the Appendix 1.C.

1.2.3 Risk Preferences and Gender Stereotypes - Part II

In the second part of the experiment (Part II), we elicit Risk Preferences and Gender Stereotypes of participants. We elicit Risk Preferences by using the Bomb Risk Elicitation Task (BRET), by Crosetto & Filippin (2013) and available on oTree[9]. In this game, there are 100 boxes with one containing a bomb. Subjects know that the bomb can be anywhere with the same probability. They simply have to choose how many boxes they want to collect. If they fall on the bomb they win a total of 0.00 €; otherwise they win 0.05 € per box collected. We elicit gender stereotypes through the Gender Role Attitudes Scale (GRAS) (García-Cueto *et al.*, 2015). This survey includes questions that elicit Transcendence and questions that elicit Sexism. Sexism is defined as discriminatory attitudes or beliefs about gender. Transcendence is characterized by the breaking of gender norms (i.e. "defenders of equality" as defined by García-Cueto *et al.*, 2015).

We also ask our subjects about the presumed topic of our study in order to rule out possible experimenter demand effect. Finally, in order to make sure that the gender of the opponent is salient enough, we ask participants to report the name chosen by their opponent during the experiment and the gender of their opponent. We conceal these questions by also asking two additional random questions (i.e., "why did you choose the tournament in Stage 3?" and "what was your final payment?").

[9] Holzmeister & Pfurtscheller (2016)

1.2.4 Procedure

The experiment was pre-registered at AEA RCT Registry[10], and approved by the Ethics Committee of the University of Paris Nanterre. We conducted the experimental sessions at the "Laboratoire d'Economie Expérimentale de la Défense" (Courbevoie, France) during the period from December 2022 to January 2023. The experiment was computerized using the oTree software (Chen *et al.*, 2016).

The experimental sessions had an approximate duration of 45 min, and the average final payment was 13.70 €. For the final payment, we used the Random Incentive System. Subjects were aware that only one out of the four stages of the game was taken into account for the final payment. The final amount payed out to each subject was the sum of the following: the payoff of the randomly chosen stage from Part I, the payoff obtained in the BRET task (Part II), and the show-up fee of 9.00 €.

1.3 Hypotheses

In this study we investigate individual behavior in a competitive environment. In particular, we focus on the individuals' reaction to a previous competition's outcome (winning or losing), and whether this reaction is influenced by the gender of the opponent. The reaction is assessed by the identification of a subject's switching point for re-entering into the competition, which depends on the relative point difference between the subject themselves and the opponent in the previous stage (see Figure 1.2). A higher switching point indicates a higher inclination to repeat the competition. We formulate hypotheses related to how individual switching point is affected by the outcome of the previous tournament and by the gender of their opponent separately for men and women.

[10]Registry number AEARCTR-0010102 (link: https://www.socialscienceregistry.org/trials/10102).

Buser *et al.* (2023) reveal that willingness to compete is influenced by the gender of the opponent: women tend to avoid competing against men, while men, anticipating this behavior, show a greater willingness to compete against women. Buser (2016) finds that men respond to losing by selecting more challenging targets. Similarly, Buser & Yuan (2019) report that men are more likely than women to re-enter competition after a loss. Moreover, men tend to be more overconfident than women, especially if the task is stereotypically masculine, such as mathematical tasks (Nosek *et al.*, 2002; Guiso *et al.*, 2008; Bordalo *et al.*, 2019). According to Bordalo *et al.* (2019), beliefs about others are similar to beliefs about own performance. On the basis of this finding and the fact that men perform better in mathematical tasks (Nosek *et al.*, 2002; Guiso *et al.*, 2008), we believe that male subjects in our sample are confident about their performance and that of their male counterparts. Based on the aforementioned stereotypes, men expect women to perform worse in mathematics and consequently men are more likely to expect to win against a female opponent. Thus, we expect men's switching point after winning to be higher against a woman than against a man (Price, 2008). On the other hand, losing against a female opponent changes expectations leading to a lower switching point because men have an "additional disutility" after losing against a high-performing woman (Datta Gupta *et al.*, 2013). We believe that this "disutility" lies in the belief that women perform poorly at mathematical tasks. This stereotype, as mentioned above, implies that men are more confident when competing against a female opponent. Therefore, negative feedback (i.e., learning that they lost) impedes men's willingness to recompete due to disappointment (as evidenced by Gill & Prowse, 2012). After losing, we predict that men's switching point is lower when they compete against women, compared to when they compete against men. This predicted outcome is similar to the one found by Datta Gupta *et al.* (2013): men refuse to compete in anticipation of the cost of losing to a woman. The notable difference in our study is that men choose to compete (or not) after being already defeated by a woman. We derive then our first hypothesis:

Hypothesis 1: Men's switching point is higher after winning against a woman than against a man. Men's switching point is lower after losing against a woman than against a

man.

Empirical studies have demonstrated that women are more likely than men to withdraw after encountering setbacks in fields dominated by males (Katz *et al.*, 2006; Goldin, 2015; Fischer, 2017). Buser (2016) finds that women tend to lower their performance after loosing. Similarly, Gill & Prowse (2014) examine the effects of winning and losing on subsequent performance, discovering that women's performance decreases following a loss. Additionally, Buser & Yuan (2019) show that women are less inclined than men to re-enter competition after experiencing a loss. Fang *et al.* (2021) find that girls are more likely to withdraw after failures in previous competitions, particularly in countries with wider gender gaps in the labor market. Regardless of losing or wining, female players are more reluctant to compete against men (Geraldes, 2020). Therefore, we conjecture that also women's switching point depends on the gender of their opponent, and not only on the outcome of their previous performance. Furthermore, since women are generally more risk averse and underconfident than men, we expect women to shy away from competition after losing. Women's decision to compete is positively influenced by previously winning in Stage 3T because of the increase in their confidence and because the feedback helps to reduce uncertainty. However, we believe that this pattern is only observed when competing against female counterparts. Against a man, women may refuse to compete even if they won previously. With larger point gaps in their favor, women choose the piece rate scheme against men because of self-restraint or the female modesty effect (Heatherington *et al.*, 1993).

Hypothesis 2: Women's switching point is lower when competing with a man compared to when competing with a woman. Additionally women's switching point is lower after loosing than after winning, regardless of the gender of the opponent.

1.4 Results

Our sample is composed of 290 participants: 58.28% of them are females, 41.38% are males, and 0.34% of the participants do not report their gender.[11] Subjects are, on average, 35.34 years old. Table 1.1 presents a summary of descriptive statistics. It includes information on competitiveness, performance in the different stages of the game, percentage of winners, risk preferences, gender stereotypes, age, and education. We show the statistics for the full sample, splitting by gender (male/female), and by condition (MM, FF, MF-FM). The variable competitiveness corresponds to the choice of payment-scheme at Stage 3 (0 = piece rate and 1 = tournament). The variables Sexism and Transcendence correspond to the sum of the Likert scale score. A lower score in Sexism corresponds to milder sexist beliefs although a higher score in Transcendence is associated with a lower compliance with sexist norms.

Subjects collect an average of 43.20 boxes out of 100 in the BRET. Women are significantly more risk-averse than men (Mann-Whitney, henceforth MWU, $p = 0.00$). Men show a significantly higher level of sexism than women (MWU, $p = 0.00$) and a lower level of transcendence (MWU, $p = 0.04$).

Overall, we do not find gender differences in performance in the task (4.88 vs. 4.37) (MWU, $p = 0.09$).[12] Under competition, women solve slightly fewer problems than men (Stage 3T and Stage 4aT; respectively MWU, $p = 0.05$ and $p = 0.02$). Nevertheless, there is an equal distribution of male and female winners in each tournament. Indeed, we find no significant differences between men and women in the proportion of winning, in any tournament (Stage 2, Stage 3T, Stage 4aT, Stage 4bT, respectively χ^2, $p = 0.60$; $p = 0.53$; $p = 0.44$; $p = 0.88$). We do not find significant differences in performance in Stage 1 (4.12 vs. 3.60, MWU, $p = 0.11$) and in Stage 2 (4.77 vs. 4.34, MWU, $p = 0.28$).

[11] Only one participant answered "Others" and we excluded this observation from the analysis. Nevertheless, we kept the gender information communicated (through their chosen name) to their opponent.

[12] See Figure 1.B1 (Appendix 1.B) for the performance evolution across the stages, based on the chosen payment scheme.

Table 1.1: Summary of descriptive statistics

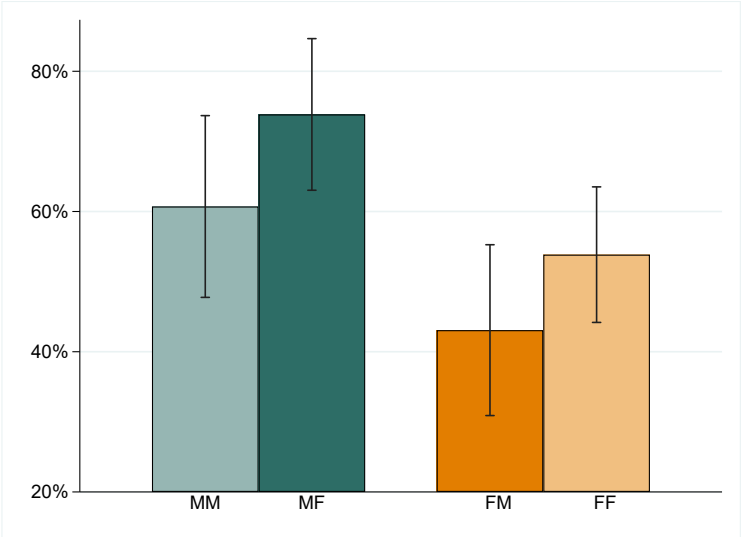
	Gender			Condition			
	Female	Male	*p*	All	MM	FF	MF-FM
Competitiveness (%)	49.70	67.75	*0.00*	57.24	60.71	53.85	58.46
Average performance	4.37	4.88	*0.09*	4.61	4.75	4.35	4.71
(s.d.)	(2.10)	(2.58)		(2.33)	(2.58)	(2.19)	(2.32)
Stage 1	3.60	4.12	*0.11*	3.84	4.07	3.65	3.89
Stage 2	4.34	4.77	*0.28*	4.53	4.66	4.30	4.64
Stage 3PR	4.34	3.82	*0.14*	4.18	4.14	3.96	4.39
Stage 3T	5.21	5.83	*0.06*	5.56	5.65	5.23	5.67
Stage 4aT	5.20	6.48	*0.02*	5.91	6.33	5.24	5.98
Stage 4aPR	4.75	5.56	*0.27*	5.15	5.46	5.30	4.74
Stage 4bT	4.64	4.53	*0.94*	4.61	4.55	4.23	4.97
Stage 4bPR	4.30	3.59	*0.09*	4.00	3.69	4.22	4.00
Distribution of winners (%)							
Stage 2	56.94	43.05	*0.60*	-	-	-	-
Stage 3T	49.01	50.98	*0.54*	-	-	-	-
Stage 4aT	48.78	51.21	*0.44*	-	-	-	-
Stage 4bT	77.14	22.86	*0.88*	-	-	-	-
Risk preferences							
Number of boxes (BRET)	38.70	49.85	*0.00*	43.20	51.60	39.03	43.20
Gender stereotypes							
Sexism	26.75	32.63	*0.00*	29.15	32.77	27.32	29.15
Transcendence	26.35	25.12	*0.04*	25.85	25.62	26.10	25.76
Age	37.82	31.88	*0.02*	35.34	30.84	39.12	34.10
Education (%)							
- Arts, Letters and Languages	10.65	6.67	*0.24*	8.97	8.93	15.38	3.88
- Law, Politics, Economics	37.28	41.67	*0.45*	38.97	35.71	28.85	48.84
- Social and Human Sciences	13.02	13.33	*0.94*	13.45	16.07	14.42	10.85
- Techno, Health and Engineering	10.65	23.33	*0.00*	15.86	28.57	11.54	13.95
- Other	28.40	15.00	*0.01*	22.76	10.71	29.81	22.48
N	169	120		290	56	104	129
	(58.28 %)	(41.38 %)			(19.38 %)	(35.99 %)	(44.64 %)

Notes: Standard deviations are in parentheses. The p-value represents the equality of distribution of the binary variables (using a χ^2 test) and continuous variable (using a MWU).

1.4.1 First entry into competition: Stage 3

In Stage 3, 49.70% of women and 67.75% of men choose the tournament payment scheme (χ^2, $p = 0.00$). This finding is in line with results from previous studies that report gender differences in competitiveness (Niederle & Vesterlund, 2007; Balafoutas & Sutter, 2012; Van Veldhuizen, 2022b). Figure 1.3 illustrates the percentage of subjects who decide to enter into the competition in Stage 3, across conditions. On the left, we present the bars representing males (MM and MF conditions), while on the right, we depict the bars for females (FM and FF conditions). We observe a non-significant pattern where both men and women are slightly more willing to compete against female opponents (χ^2, respectively $p = 0.14$ and $p = 0.17$). This result aligns with findings from previous studies (Datta Gupta *et al.*, 2013; Burow *et al.*, 2017). The elicitation of beliefs that we conduct after the task shows that men are significantly more confident than women: 75.21% of men expect to win versus 55.03% of women (χ^2, $p = 0.00$). For both men and women, confidence is not affected by the gender of the opponent (χ^2, respectively $p = 0.20$ and $p = 0.23$).

Figure 1.3: Percentage of subjects choosing the competition in Stage 3, by gender and condition



Notes: Confidence intervals at 95%.

In Table 1.2, we conduct four specifications using a Probit model to understand the driving factors of deciding on the choice of payment-scheme in Stage 3. The dependent variable takes the value 1 if the subject chooses the tournament, and 0 otherwise. Columns 1 and 2 refer to male observations, and columns 3 and 4 to female observations. In all specifications the explanatory variables are: Female opponent (takes value 1 if the opponent is a female, and 0 otherwise), Risk aversion (takes value 1 if the subject collects less than 50 boxes at the BRET task, and 0 otherwise), Age, Sexism, Transcendence, and *Beliefs 2* (takes value 1 when the participant expects to win in Stage 2, and 0 otherwise). In columns 2 and 4, we add the interaction between the gender of the opponent and the beliefs on Stage 2 (Female opponent $\times$ *Beliefs 2*).

Entry into the competition seems to be mainly influenced by *Beliefs 2*: men and women who expect to win in Stage 2 are significantly more likely to choose the tournament payment scheme than the piece rate in Stage 3. We observe a significant and positive effect of Transcendence for males (columns 1 and 2), suggesting that lower gender stereotypes (i.e. being more transcendent) significantly increase the probability to choose the tournament in Stage 3. We do not find any significant coefficient for Female opponent[13], Risk aversion[14], Age, or Sexism on the decision to compete. Our interaction term is insignificant for males and females, indicating that confident participants are indifferent to the gender of their opponent when deciding whether to compete in Stage 3. In fact, the main driver to decide whether to compete or not are beliefs, meaning that participants who expect to win in Stage 2 are more likely to enter into competition in Stage 3, irrespective of their own gender or that of their opponents.

[13]Even if the gender of the opponent does not seem to have, on average, an impact on the decision to compete, two subjects reported that they took this variable into account when choosing the tournament. Indeed, to the question "why did you choose the tournament?", the participants' answer was: "my opponent's name was female" and "because I thought I had a good chance of winning against a girl, girls are generally less good at maths than boys".

[14]On average, competitive participants collect 44.22 boxes, and non-competitive collect 41.84 boxes (MWU, $p = 0.25$).

Table 1.2: Determinants of entry into competition in Stage 3.

	Entry competition			
	Males		Females	
	(1)	(2)	(3)	(4)
Female opponent	0.40	0.71	0.24	0.27
	(1.50)	(1.36)	(1.14)	(0.88)
Risk aversion	0.22	0.22	0.01	0.01
	(0.84)	(0.82)	(0.04)	(0.04)
Age	0.00	0.00	-0.00	-0.00
	(0.21)	(0.18)	(-0.69)	(-0.67)
Sexism	0.00	0.00	-0.01	-0.01
	(0.00)	(0.08)	(-0.57)	(-0.57)
Transcendence	0.10**	0.10**	-0.01	-0.01
	(2.66)	(2.69)	(-0.27)	(-0.27)
Beliefs 2	1.04***	1.25**	0.78***	0.82*
	(3.48)	(2.95)	(3.79)	(2.47)
Female opponent X *Beliefs 2*		-0.44		-0.06
		(-0.73)		(-0.15)
Constant	-3.05*	-3.23*	0.01	-0.02
	(-2.31)	(-2.37)	(0.01)	(-0.01)
Observations	120	120	169	169

Notes: Probit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-4: Entry competition. Specifications 1-2: Only males. Specifications 3-4: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

1.4.2 Second entry into competition: Stage 4a

In this section, the analysis focuses on the second choice of payment-scheme, i.e., choice of tournament at Stage 4a.[15] In this stage, we measure subjects' competitiveness based on their switching point. The switching point is defined as the performance difference at which individuals opt to switch from engaging in competition to refraining, or conversely. The advantage of the switching point is that it provides a continuous measure of subjects' level of competitiveness. After losing, participants switch from accepting to compete for smaller points gaps to refusing to compete for larger points gaps.[16] Reversely, after winning, the switching point goes in the opposite direction: participants switch from refusing to accepting.[17] To facilitate the interpretation of the results, we standardize the switching point to ensure that it points always in the same direction. Therefore, the higher the value of the switching point, the more willing is a participant to compete again.

Figure 1.4 illustrates the distribution of the switching points, conditional on the outcome of Stage 3T (winning, in panel (a), and losing, in panel (b)) and the gender of the participant (green bars for males, and orange bars for females). Overall, we observe similar patterns in the distribution of the switching points for men and women (Two-sample Kolmogorov–Smirnov, $p = 0.81$). After winning (Figure 1.4a), the distribution is skewed to the right with a peak in the upper limit (i.e., a switching point of 11), while after losing (Figure 1.4b), the distribution is skewed to the left with a mild peak in a switching point of 2 for men and 3 for women. After losing, men and women are equally willing to accept the competition

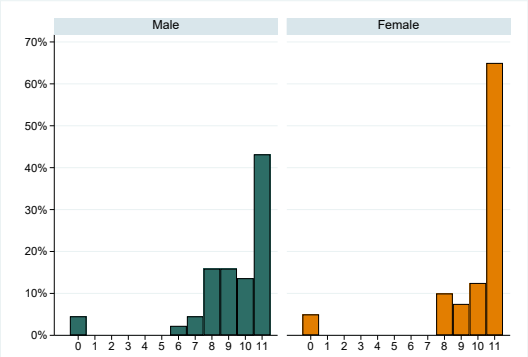
[15] In Appendix 1.C, we analyze the data corresponding to the subjects that choose the piece rate payment scheme in Stage 3. Moreover, Appendix 1.D includes additional analyses.

[16] For example, after losing, a switching point of 8 means that the subject accepts to compete for a point gap inferior at 8. If the opponent solved 8 problems (or more) than the subject, the subject refuses to compete. This approach to coding the variable excludes by default participants who have switched more than once although our results are robust when including them. Table 1.E1 in Appendix 1.E presents four specifications using a Tobit model with switching point as dependent variable ranging from 0 to 11. In this robustness check, the switching point of those who switch more than once (7 men and 7 women) are taking into account. However, we always exclude from the analysis those who only accept to compete for one point gap (N = 2) and those who switch in the opposite direction for the switching point after winning (N = 15).

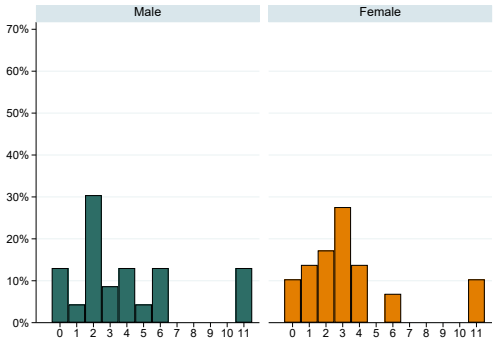
[17] For example, after winning, a switching point of 5 means that the participant does not want to compete for point gaps inferiors (or equal) to 5. We code the switching point after winning following the same logic as after losing. Only a minority of the subjects (N = 15) switch from accepting to refusing to compete.

Figure 1.4: Distribution of the switching point, by outcome in Stage 3T and gender of the subject.

(a) After winning



(b) After losing

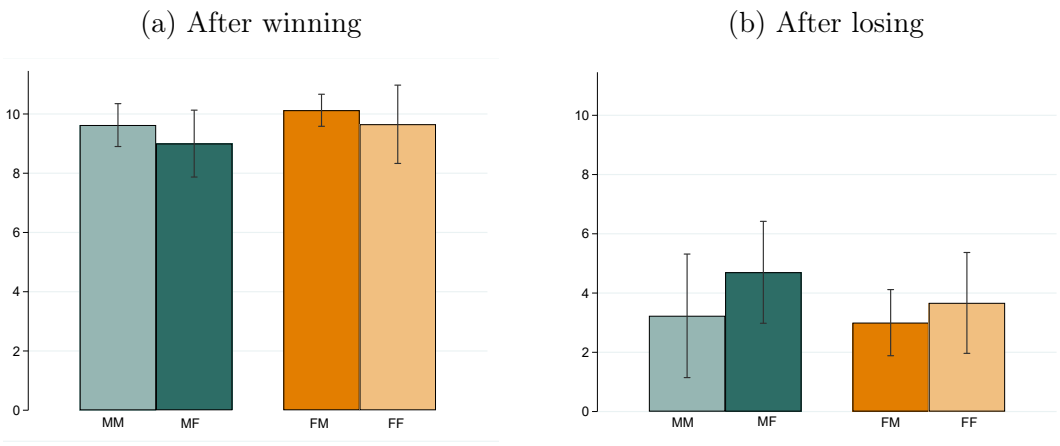


for any points gap, i.e., a switching point equals to 11 (13.04% vs. 10.34%, χ^2, $p = 0.51$). Nevertheless, after winning, women are more likely to compete again for any points gap, compared to men (65.00% vs. 43.18%, χ^2, $p = 0.04$). The findings suggest that both men and women exhibit increased levels of competitiveness after winning, with this trend being particularly pronounced among women.

We further investigate whether the gender of the opponent affects the decision to compete again. Figure 1.5 shows the average switching point for men and women, conditional on the outcome of Stage 3T (winning or losing) and the condition they were allocated to (MM, FF, MF and FM). Overall, we find no significant differences in the average switching point between men and women (respectively 7.36 vs. 7.10, MWU, $p = 0.83$).[18]After losing, both men and women are slightly more willing to play against females, although the difference is not significant. The average switching point in MM is 3.23 and 4.70 in MF (MWU, $p = 0.06$). Defeated females' switching point is also slightly (but not significantly) higher in FF than in FM (respectively 3.67 vs. 3.00, MWU, $p = 0.94$). However, we find the opposite pattern after winning: both men and women accept slightly (but not significantly) more to compete against a defeated male player. The average switching point in MM is 9.62 vs. 9.04

[18]The decline of gender differences in competition could be explained by the absence of gender differences in confidence after Stage 3T. When we elicit their beliefs, men and women equally expect to win in Stage 4aT (χ^2, $p = 0.19$).

Figure 1.5: Average switching point, by outcome in Stage 3T and condition



Notes: Confidence intervals at 95%.

in MF (MWU, $p = 0.95$). The average switching point in FM is 10.12 vs. 9.71 in FF (MWU, $p = 0.25$).

In Table 1.3, we explore the determinants of the switching point. We conduct Tobit estimations with switching point as dependent variable ranging from 0 to 11. Our explanatory variables are: Female opponent, Risk aversion, Age, Sexism, Transcendence. In specifications 1-2 and 5-6 we include the variable Winner Stage 3T which takes value 1 if the subject wins in Stage 3, and 0 otherwise. We also include in specifications 3-4 and 7-8 the variable Overconfidence Stage 3T that takes value 1 when a participant expects to win at Stage 3T but actually loses. It takes value 0 in either of the three following cases: 1-expects to win and actually wins, 2-expects to loose and actually loses, or 3-expects to loose and actually wins. In columns 2 and 6, we add the interaction term: Female opponent $\times$ Winner Stage 3T. In columns 4 and 8 we include the interaction between the gender of the opponent and the overconfidence in Stage 3T (Female opponent $\times$ Overconfidence Stage 3T).[19]

[19]We test the robustness of our results to a specification that includes a variable for confidence instead of overconfidence. This variable takes the value 1 if the subject is overconfident, 0 if the beliefs are accurate, and -1 if the subject is underconfident. Results of our Tobit estimations on the switching point incorporating this new measure as an explanatory variable are presented in Table 1.E2 (Appendix 1.E). Furthermore, we conduct an additional robustness check by removing subjects who do not correctly report the gender of their opponent at the end of the experiment. Table 1.E3 in Appendix 1.E presents results of Tobit estimations on the switching point of those who correctly guess the gender of their opponent. We perform this robustness

In line with our descriptive results, we find a significant and positive coefficient of Winner Stage 3T in our specifications 1-2 and 5-6 indicating that winning in the previous competition (in Stage 3) significantly increases the switching point of both males and females. In our second specification, the interaction term Female opponent × Winner Stage 3T has a negative and significant effect, indicating that male winners lower significantly their switching point when winning against a female compared to when winning against a male. In other words, after winning, men have a lower willingness to compete again against a woman than against a man. The interaction term is not significant for females in column 6, indicating that, after winning, women are not influenced by the gender of their opponent when deciding whether to repeat the competition.

We find a significant and negative coefficient of Overconfidence Stage 3T in our four specifications where the variable Overconfidence Stage 3T is present (columns 3-4 and 7-8), meaning that when participants are disappointed, i.e. they expect to win but end up losing, they tend to lower their switching point. This result goes in line with Gill & Prowse (2012)'s findings on disappointment aversion. In other words, participants are discouraged to repeat the competition when they learn that their outcome is not the one expected. In column 4, a positive and statistically significant effect emerges within the interaction term Female opponent × Overconfidence Stage 3T for males, indicating that the switching point is significantly higher when they are disappointed against women compared to when they are disappointed against men. In other words, the disappointment effect is less intense when a male engages in competition with a female as opposed to a male counterpart. This observation implies that, disappointed men are more eager to repeat the competition against a woman than against a man. Again, the interaction term Female opponent × Overconfidence Stage 3T is not significant for women, indicating that women do not make difference between having a male or a female opponent.

To summarize our findings, we observe a deviation from Hypothesis 1 in men's reactions after winning. After losing, men exhibit a behavioral shift, contrary to our hypothesis. Win-

check to ensure that our results are valid for those who identify which gender they are competing against.

Table 1.3: Determinants of Switching point

	Switching point							
	Males				Females			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female opponent	0.31	2.42	0.14	-0.66	-0.13	0.92	0.13	-0.10
	(0.43)	(1.93)	(0.18)	(-0.76)	(-0.19)	(0.83)	(0.16)	(-0.11)
Risk aversion	-0.92	-0.98	-1.06	-0.96	-0.42	-0.53	-0.16	-0.19
	(-1.23)	(-1.35)	(-1.33)	(-1.23)	(-0.58)	(-0.74)	(-0.20)	(-0.24)
Age	-0.04	-0.04	-0.06*	-0.06*	0.03	0.04	0.02	0.02
	(-1.61)	(-1.64)	(-2.05)	(-2.15)	(1.50)	(1.77)	(0.79)	(0.87)
Sexism	0.03	0.06	0.07	0.09	0.05	0.04	0.06	0.06
	(0.78)	(1.40)	(1.48)	(1.99)	(1.06)	(0.85)	(1.02)	(0.99)
Transcendence	0.01	0.06	0.10	0.16	0.16	0.17	0.15	0.16
	(0.08)	(0.56)	(0.82)	(1.28)	(1.12)	(1.26)	(0.95)	(1.00)
Winner Stage 3T	5.31***	7.02***			6.30***	7.37***		
	(6.79)	(6.21)			(9.10)	(6.64)		
Female opponent × Winner Stage 3T		-3.25*				-1.79		
		(-2.06)				(-1.22)		
Overconfidence Stage 3T			-6.10***	-7.98***			-6.06***	-6.51***
			(-5.74)	(-5.83)			(-7.23)	(-4.84)
Female opponent × Overconfidence Stage 3T				4.53*				0.73
				(2.14)				(0.42)
Constant	4.05	0.78	5.67	3.89	-2.87	-3.89	2.92	2.81
	(0.95)	(0.17)	(1.24)	(0.86)	(-0.65)	(-0.87)	(0.57)	(0.55)
Observations	66	66	66	66	69	69	69	69

Notes: Tobit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-8: Switching point. Specifications 1-4: Only males. Specifications 5-8: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

ning against a female opponent decreases the switching point. Additionally, when we consider only overconfident men, we observe that disappointment diminishes the inclination to compete against a male opponent. On the other hand, overconfident and hence disappointed men, when outperformed by a female, appear more inclined to re-enter the competition. Therefore, we reject our first hypothesis and we present our first key result:

Result 1: *After winning against women, men exhibit a decrease in their switching point, whereas disappointment for losing against women (i.e. when men expect to win but they lose) leads to an increase in their switching point.*

In contrast to our second hypothesis, the gender of the opponent does not significantly influence women's switching points. Remarkably, women do not demonstrate increased hesitancy to re-enter competition when facing men as opposed to women. One plausible explanation is that non-competitive women, who opt for the piece rate in Stage 3 and are subsequently excluded from our sample, might be the ones avoiding competition against men. This nuanced perspective aligns with the insights from Geraldes (2020), suggesting that women's reluctance to compete might be more directed towards competing against men specifically, rather than a general aversion to competition.[20] We confirm the second part of our hypothesis, since women's (and men's) switching point is lower after loosing than after winning, regardless of the gender of their opponent. This leads us to formulate our second key result:

Result 2: *After winning women are more willing to repeat a competition. Disappointed women (i.e. when they expect to win but they lose) are less likely to repeat the competition. However, on average, women exhibit indifference to the gender of the opponent when deciding to re-enter competition.*

1.4.3 Other analysis: Performance

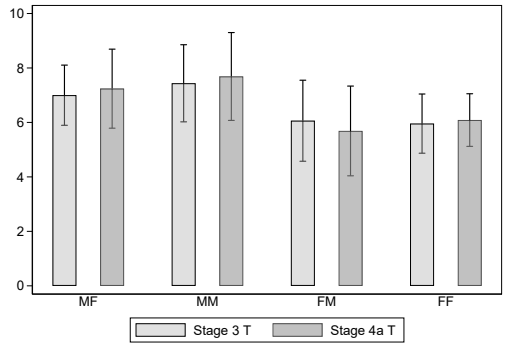
We are further interested in analyzing the performance in the tournament and its evolution over the stages. Men perform significantly better in Stage 4aT after losing in Stage 3T against a woman (MF) than against a man (MM). As shown in Figure 1.6 panel (b), men significantly improve their performance between Stage 3T and Stage 4aT against a female opponent (Wilcoxon signed-rank, $p = 0.02$): they solve on average 2.00 more problems in Stage 4aT, compared to Stage 3T. When re-facing a male opponent after losing (MM), men solve slightly (but not significantly) fewer problems (i.e., difference of -0.80; Wilcoxon signed-

[20] We test the robustness of our results by replicating Table 1.3 in Table 1.E4 (Appendix 1.E) using OLS estimations and we find qualitatively the same results.

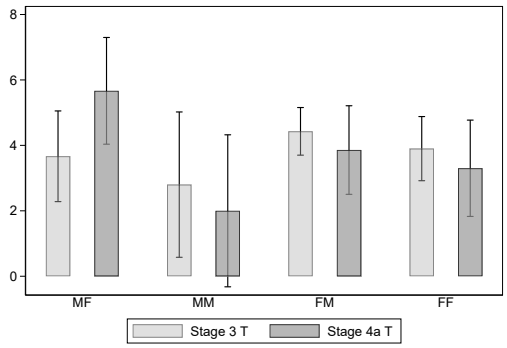
rank, $p = 0.25$). After winning, differences in performance are not significant (panel (a)). We do not find any significant results in women's performance.

Figure 1.6: Performance in Stage 3T and Stage 4aT, by outcome in Stage 3T and condition

(a) Winners in Stage 3T



(b) Losers in Stage 3T



Notes: Confidence intervals at 95%.

In Table 1.4, we run an OLS estimation with performance in Stage 4aT as dependent variable. We break down our sample and we include only males in columns 1 and 2 and only females in columns 3 and 4. The explanatory variables included in the models are: Female opponent, Loser Stage 3T (=1 if the participant lost in Stage 3T, 0 otherwise), and the interaction term Female opponent × Loser Stage 3T, and Performance in Stage 3T. In columns 2 and 4, we add the following control variables: Age, Sexism, and Transcendence. We observe a positive and significant coefficient for Performance in Stage 3T, indicating that the higher the performance in Stage 3T, the higher the performance in Stage 4aT. The interaction term in columns 1 and 2 also exhibits a positive and significant coefficient, confirming that, after losing, men enhance their performance against a female opponent compared to a male opponent. However, a similar effect is not observed for women. These results remain robust even with the inclusion of control variables.

Consistent with Result 1, we observe that after losing, men exhibit an increased determination to succeed against female opponents. This is evident in two dimensions: a higher switching point and an increase in performance after losing against women.

Table 1.4: Determinants of performance in Stage 4aT

	Performance in Stage 4aT			
	Males		Females	
	(1)	(2)	(3)	(4)
Female opponent	-0.12	-0.35	0.31	0.57
	(-0.19)	(-0.52)	(0.58)	(1.00)
Age		-0.01		-0.02
		(-0.30)		(-1.48)
Sexism		0.06		0.03
		(1.76)		(0.93)
Transcendence		0.12		0.11
		(1.14)		(1.12)
Loser Stage 3T	-1.48	-2.57*	-0.50	-0.33
	(-1.89)	(-2.58)	(-0.78)	(-0.50)
Female opponent × Loser Stage 3T	3.12**	4.61***	-0.11	-0.39
	(3.40)	(3.82)	(-0.12)	(-0.42)
Performance in Stage 3T	0.91***	0.92***	0.81***	0.80***
	(7.52)	(7.89)	(7.43)	(6.31)
Constant	0.93	-3.67	0.75	-2.46
	(0.96)	(-1.00)	(0.97)	(-0.74)
Observations	54	54	56	56

Notes: OLS model regressions.t statistics in parentheses and standard errors are robust. Dependent variable in 1-4: Performance in Stage 4aT. Specifications 1-2: Only males. Specifications 3-4: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

1.5 Conclusions

We conduct an experiment in which participants have to perform in a simple task consisting of adding two digits. They face two main decisions: First, choosing between being paid for the task based on a piece rate or a tournament. After observing the outcome (winning or losing), they made a decision on payment scheme for repeating the task. Our primary interest lies in understanding people's preferences for competition and their reactions after winning versus losing. Specifically, we aim to examine whether individuals are more likely to repeat the competition or refrain from it, focusing on how the preference for competition is influenced by the prior competition's outcome, and if the gender of the opponent plays a role in this decision-making process. Additionally, we introduce a novel design that obtains a continuous measure of subjects' willingness to compete, deviating from the standard binary decision framework.

Our data shows that in the first choice, men are more willing to compete than women, consistent with previous studies. Participants, however, show indifference to the gender of their opponent, and their decision to compete is mainly based on their expectation to win in previous stages where the task is performed. In the second choice to compete (or not), regardless of their opponent's gender, both men and women exhibit increased levels of competitiveness after winning, with this trend being particularly pronounced among women. Nevertheless, we observe that men decrease their willingness to repeat the competition after winning against women. One plausible explanation for these findings could be that men may previously have built a reputation for prevailing over women. Consequently, they may be inclined to avoid the potential risk of losing a competition against a woman, as doing so could jeopardize their established reputation. In essence, men might be motivated to uphold their perceived dominance, thereby preventing women from outperforming them and hence avoiding potential disutility (He & Noussair, 2021).

We further observe that overconfidence is a crucial driver in the choice to avoid competition for a second time. Overall, participants compete less when they are disappointed

with their outcome, i.e., they expect to win and end up loosing. Males are less affected by disappointment when they loose against women, and they are more eager to repeat the competition. In line with this, we find significant results regarding the reaction after losing, in terms of performance. Men's performance significantly improves after losing against females, revealing that males have a significantly higher drive to win over their female opponents compared to male opponents. This result suggests that the losers are indeed stimulated by the challenger option, i.e., winning over a woman who has previously outperformed them (Buser, 2016). Men's willingness to outperform their female opponent in the next round might be driven by possible factors such as shame, revenge, or pride. On the other hand, women are not affected in any case by the gender of their opponent: they are more likely to compete after winning and less likely to compete when they are disappointed, regardless of their opponent's gender. Overall, men seem to be reluctant to be outperformed by women which can explain the underrepresentation of women in senior and high-paying positions (De Paola & Scoppa, 2015).

This work contributes substantially to the existing literature and provides insightful findings to better understand gender differences in competition. We confirm previous findings showing that men are generally more sensitive to feedback and the gender of the opponent, compared to women. For example, Dilmaghani (2022) find that in super-elite chess games, males are quicker to quit when facing a female opponent compared to a male, while female players exhibit consistent behavior regardless of the opponent's gender. Furthermore, Berlin & Dargnies (2016) indicate that women are more focused on their performance level, while men are more influenced by the skill level of their competitors. Additionally, Wozniak *et al.* (2009) show that feedback diminishes gender differences by reducing men's reluctance to participate. Finally, Baier *et al.* (2018) conclude that men are more responsive to feedback in task choice, whereas women display greater stability in task selection across different information.

Our experiment concludes the game at Stage 4 without providing participants with information regarding the decisions or performances of their opponents in Stage 4a. This feature

could serve as promising for future research, offering the opportunity to investigate potential renunciation effects when such information is revealed. Additionally, it could provide further evidence regarding whether high-performing women exert less effort when facing a persistent male opponent.

1.A Perception of Names

Table 1.A1: Answers to the question: "How do you perceive these names?" (N=112)

	Adam	**Arthur**	**Louis**	**Lucas**	**Raphaël**	Gabriel	Léo	Jules	Maël	Noah
Male	**99.11**%	**99.11**%	**98.21**%	**100.00**%	**99.11**%	96.43%	97.32%	98.21%	70.54%	91.07%
Female	0.89%	0.89%	1.29%	0.00%	0.89%	3.57%	2.68%	1.79%	29.46%	8.93%

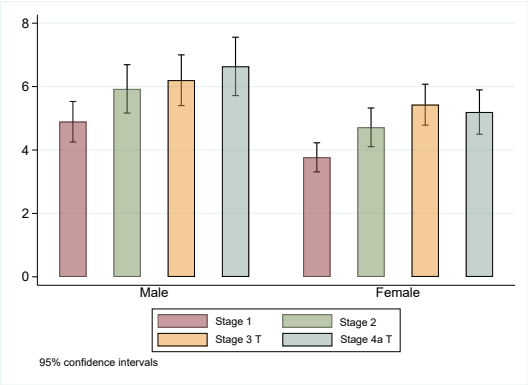
	Alice	**Anne**	**Louise**	**Mia**	**Rose**	Jade	Emma	Ambre	Alba	Romy
Male	0.00%	0.89%	0.89%	0.00%	0.00%	4.47%	1.79%	1.79%	6.25%	16.07%
Female	**100.00**%	**99.11**%	**99.11**%	**100.00**%	**100.00**%	95.53%	98.21%	98.21%	93.75%	83.93%

Notes: In bold, the names chosen for our experiment: Adam, Arthur, Louis, Lucas and Raphaël for males; Alice, Anne, Louise, Mia and Rose for females.

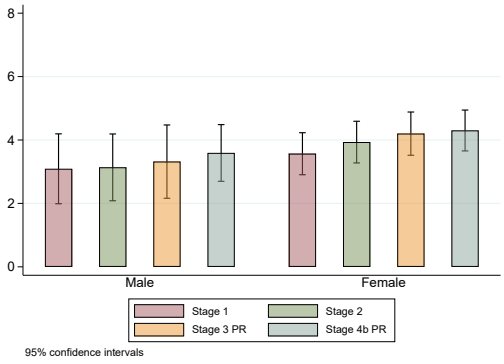
1.B Performance across Stages

Figure 1.B1: Average performance in each stage of the game, by gender and conditional on the decision to compete made at Stage 3

(a) Competitors



(b) Non competitors



1.C Stage 4b: WTP to Avoid Competition

Subjects who choose the piece rate in Stage 3 are assigned to Stage 4b, in which we elicit their willingness to pay to avoid being forced to compete. From the data collected in Stage 4b, subjects show a relatively low aversion to compete. Indeed, 44.35% choose a *WTP* inferior to 0.75 € (the maximum amount being 2.00 €). As shown in Figure 1.C1, women's and men's distribution of *WTP* is similar (Two-sample Kolmogorov-Smirnov, $p = 0.24$). In line with the literature, we previously demonstrate that women are less competitive than men in Stage 3. Although, when offered the opportunity to reconsider their choice, women do not manifest a strong aversion, compared to men. Their *WTP* is not significantly different (MWU, $p = 0.07$).

Figure 1.C1: Distribution of the WTP at Stage 4b, by gender

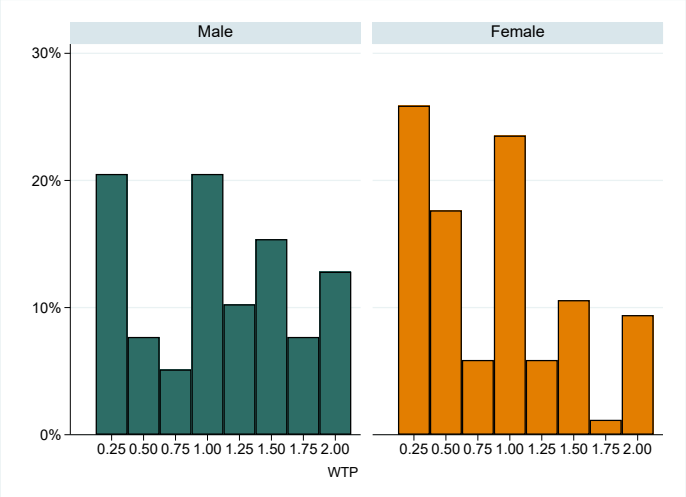


Figure 1.C2 presents the evolution of performance from Stage 3PR and Stage 4bT by condition. Both men and women perform slightly better against females than males in Stage 4b (compared to Stage 3PR), but this difference is not significant (Wilcoxon signed-rank, respectively $p = 0.66$ for MF; $p = 0.69$ for MM; $p = 0.14$ for FF; $p = 0.70$ for FM). In Stage 4bT women solve on average 4.64 problems and men 4.53 (MWU, $p = 0.94$). Compared to Gill & Prowse (2014), women in our sample do not perform worse than men when forced to

compete.

We investigate the factors contributing to determine the *WTP* (min=0.25, max=2.00), by using a Tobit model. Table 1.C1 presents the results. Our explanatory variables are: Female opponent, Risk aversion, and *Beliefs 2*. In columns 2 and 4 we add control variables. We do not find any significant effect of the variables on the *WTP*.

Figure 1.C2: Average performance in Stage 3PR and Stage 4bT, by condition

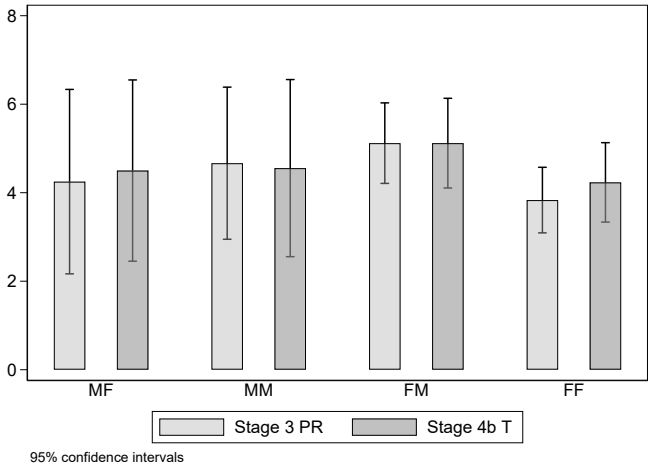


Table 1.C1: Willingness to pay to avoid competition in Stage 4b

	WTP			
	Males		Females	
	(1)	(2)	(3)	(4)
Female opponent	-0.16	-0.19	-0.04	-0.05
	(-0.79)	(-0.96)	(-0.37)	(-0.41)
Risk aversion	-0.13	-0.11	-0.05	-0.05
	(-0.63)	(-0.56)	(-0.36)	(-0.42)
Beliefs 2	-0.13	-0.19	0.22	0.23
	(-0.66)	(-0.94)	(1.78)	(1.89)
Age		0.00		-0.01*
		(0.34)		(-2.02)
Sexism		-0.00		0.01
		(-0.38)		(1.51)
Transcendence		-0.02		0.01
		(-1.14)		(0.72)
Constant	1.27***	1.96*	0.84***	0.46
	(7.13)	(2.29)	(5.80)	(0.74)
Observations	39	39	85	85

Notes: Tobit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-4: WTP in Stage 4b. Specifications 1-2: Only males. Specifications 3-4: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

1.D Additional Analyses

At the end of the experiment, subjects were asked about their opponent, the presumed topic of our study, and their motivations to choose (or not) the tournament at Stage 3. This Section presents the textual analysis of participants' responses, which complements the econometric analysis.

When asked about the presumed topic, 30.64% of participants mention "risk" or even "risk aversion" with subjects comparing the risk of choosing the tournament with the risk of running into the bomb during the BRET task (Table 1.D1). "Competition" is cited by 4.80% of the subjects and "gender" is only mentioned by 4 participants.[21]

The main reason given by participants for their choice to enter (or avoid) the tournament is confidence (Table 1.D2). Indeed, 36.90% of subjects mention confidence in their own performances (73.83%) or regarding their opponent (26.17%) when choosing the compensation scheme at Stage 3.[22] Furthermore, women integrate the idea that they are not good in mathematical tasks, as 21.89% of female participants report a lack of confidence in their calculation skills (against 7.44% of men). 32.23% of men emphasize their confidence in their abilities when asked "why did you choose the tournament" against 13.02% of women. Female participants are significantly more likely to expose their under-confidence in their response, compared to men (χ^2, $p = 0.00$). Some participants also demonstrate a payoff maximization strategy when they made their choice in Stage 3. Indeed, 15.86% of subjects mention the term "payoff". Finally, 4.83% mention their willingness to seek challenges and 4.13% refer to their competitiveness.[23]

[21] Besides, "gender" was mentioned as a measure that our study would take into account to evaluate the skills in mathematics among the French population.

[22] For example, confidence in themselves corresponds to answers such as "I am good at mental arithmetic" or "I think I'm bad at mental arithmetic". While confidence about each others' skills is reflected in responses such as "I feared that my opponent was good" or "I have a more literary profile and I didn't want to try the tournament with someone I thought was definitely stronger". We do not observe gender differences in the reference point for confidence.

[23] In this experiment, it is interesting to note that competitiveness is indeed perceived as "assertive and uncooperative" behaviors by some participants. "Altruism" is mentioned 3 times (2 are women) as an argument to not enter into the competition. It is presented as the opposite of competitiveness. For example:

Table 1.D1: Answers to the question: "What is the presumed topic of our study?"

Answers	Frequency (%)
Others	54.22
Risk	30.64
Mathematical capacities	7.24
Competition	4.80
Reaction to pressure	1.72
Gender	1.38

Table 1.D2: Answers to the question: "Why did you (not) choose the tournament at Stage 3?"

Answers	Frequency (%)
Confidence	36.90
Others	27.26
Payoff	15.86
Risk	8.96
Challenge	4.83
Competitiveness	4.13
Altruism	1.03
Fear of losing	1.03

"*I am not competitive, I prefer to help each other and encourage pedagogy.*"

1.E Robustness Checks

Table 1.E1: Switching point (including subjects who switch more than once)

	Switching point			
	Males		Females	
	(1)	(2)	(3)	(4)
Female opponent	0.13	-0.73	0.41	0.38
	(0.17)	(-0.87)	(0.53)	(0.41)
Risk aversion	-0.83	-0.71	-0.03	-0.03
	(-1.07)	(-0.94)	(-0.04)	(-0.04)
Age	-0.06*	-0.06*	0.01	0.01
	(-2.23)	(-2.28)	(0.50)	(0.50)
Sexism	0.05	0.08	0.02	0.02
	(1.07)	(1.74)	(0.50)	(0.49)
Transcendence	0.06	0.15	0.03	0.04
	(0.49)	(1.25)	(0.25)	(0.26)
Overconfidence Stage 3T	-5.98***	-8.03***	-6.08***	-6.14***
	(-6.04)	(-6.10)	(-7.51)	(-4.55)
Female opponent × Overconfidence Stage 3T		4.66*		0.09
		(2.31)		(0.06)
Constant	7.56	4.73	6.80	6.78
	(1.83)	(1.13)	(1.50)	(1.49)
Observations	73	73	76	76

Notes: Tobit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-4: Switching point. Specifications 1-2: Only males. Specifications 3-4: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

Table 1.E2: Switching point with with confidence

	Switching point			
	Males		Females	
	(1)	(2)	(3)	(4)
Female opponent	-0.15	-0.52	-0.15	-0.58
	(-0.17)	(-0.62)	(-0.18)	(-0.62)
Risk aversion	-1.59	-1.26	-0.03	-0.14
	(-1.96)	(-1.56)	(-0.03)	(-0.16)
Age	-0.05	-0.05	0.03	0.04
	(-1.92)	(-1.95)	(1.39)	(1.66)
Sexism	0.03	0.06	0.06	0.04
	(0.72)	(1.26)	(0.92)	(0.69)
Transcendence	0.01	0.09	0.24	0.27
	(0.11)	(0.67)	(1.40)	(1.60)
Confidence Stage 3T	-4.05***	-6.03***	-4.15***	-5.35***
	(-5.32)	(-4.95)	(-5.94)	(-4.48)
Female opponent × Confidence Stage 3T		3.26*		1.88
		(2.07)		(1.23)
Constant	8.60	6.24	-0.87	-1.38
	(1.86)	(1.34)	(-0.16)	(-0.26)
Observations	66	66	69	69

Notes: Tobit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-4: Switching point. Specifications 1-2: Only males. Specifications 3-4: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

Table 1.E3: Switching point (only subjects who correctly guess the gender of their opponent)

	Switching point			
	Males		Females	
	(1)	(2)	(3)	(4)
Female opponent	0.07	-1.12	0.91	1.25
	(0.08)	(-1.09)	(1.26)	(1.35)
Risk aversion	-1.49	-1.25	0.26	0.32
	(-1.63)	(-1.43)	(0.35)	(0.41)
Age	-0.04	-0.04	0.01	0.01
	(-1.43)	(-1.48)	(0.58)	(0.38)
Sexism	-0.00	0.01	-0.00	-0.00
	(-0.07)	(0.20)	(-0.05)	(-0.03)
Transcendence	-0.05	0.01	0.19	0.18
	(-0.29)	(0.05)	(1.18)	(1.12)
Overconfidence St 3T	-6.32***	-9.73***	-7.11***	-6.63***
	(-5.10)	(-5.32)	(-9.15)	(-5.87)
Female opponent × Overconfidence St 3T		6.23*		-0.90
		(2.57)		(-0.59)
Constant	11.85	10.60	3.15	3.32
	(1.94)	(1.82)	(0.60)	(0.63)
Observations	51	51	51	51

Notes: Tobit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-4: Switching point. Specifications 1-2: Only males. Specifications 3-4: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

Table 1.E4: Switching point (OLS estimation)

	Switching point							
	Males				Females			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female opponent	0.22	1.98	0.06	-0.60	-0.12	0.84	0.14	-0.05
	(0.35)	(1.62)	(0.09)	(-0.69)	(-0.21)	(0.87)	(0.20)	(-0.06)
Risk aversion	-0.72	-0.77	-0.83	-0.73	-0.31	-0.41	-0.04	-0.06
	(-0.97)	(-1.05)	(-1.05)	(-0.98)	(-0.41)	(-0.56)	(-0.05)	(-0.07)
Age	-0.04	-0.04	-0.05	-0.06	0.03	0.04	0.02	0.02
	(-1.55)	(-1.52)	(-1.79)	(-1.84)	(1.17)	(1.33)	(0.68)	(0.70)
Sexism	0.03	0.05	0.06	0.08	0.05	0.04	0.06	0.06
	(0.54)	(0.95)	(1.16)	(1.44)	(1.41)	(1.17)	(1.04)	(0.99)
Transcendence	0.00	0.05	0.09	0.14	0.13	0.15	0.12	0.13
	(0.03)	(0.35)	(0.57)	(0.83)	(1.45)	(1.53)	(0.86)	(0.86)
Winner Stage 3T	5.20***	6.62***			6.20***	7.18***		
	(6.48)	(6.58)			(8.52)	(12.29)		
Female opponent × Winner Stage 3T		-2.72				-1.65		
		(-1.88)				(-1.32)		
Overconfidence Stage 3T			-5.72***	-7.17***			-5.89***	-6.24***
			(-6.82)	(-7.96)			(-6.91)	(-7.13)
Female opponent × Overconfidence Stage 3T				3.63**				0.58
				(2.86)				(0.38)
Constant	4.32	1.62	6.18	4.83	-2.14	-3.06	3.62	3.55
	(0.76)	(0.29)	(1.10)	(0.85)	(-0.79)	(-0.99)	(0.75)	(0.72)
Observations	66	66	66	66	69	69	69	69

Notes: OLS model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-8: Switching point. Specifications 1-4: Only males. Specifications 5-8: Only females. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

1.F Instructions (translated from French)

Hello everyone, welcome to this experimental session!

Thank you for participating in this session, your decisions are very important for this research. Please find on your desk an information note, two consent forms and two payment receipts -one for you and one for the experimenter. Please fill in the consent form before the session starts; the gratification receipt will be filled in at the end of the session. The consent form specifies that you agree to share your data for research purposes; your data will surely be processed anonymously and only for research purposes.

During the session, you will be asked to perform mental arithmetic. The use of any electronic device is forbidden but you can use the piece of paper provided on your desk. Please put away your mobile phone and/or calculator. If you fail to comply with this rule, your data will be scientifically unusable, and a great loss to our research. Besides, you won't be paid.

Please read the instructions carefully. Communication with other participants is not allowed during the whole session. If you have any question, please raise your hand and the experimenter will come to your cubicle and answer you privately.

Instructions for Part 1

This experimental session is divided into two parts: Part 1 and Part 2. In Part 1, you will be asked to perform the same specific task during 4 stages. Before the start of each stage, specific instructions will be given to you. In Part 2, you will play a game and answer some questions.

Once you have completed the 4 stages of Part 1, a random draw will determine which stage

will be chosen for your payment.

Your total payment in cash at the end of the session includes: a participation fee of 9 € + the earnings obtained in Part 1 (i.e. the earnings corresponding to the randomly chosen stage) + the earnings obtained in Part 2 (that will depend on your decision and the others' in a game). Your total earnings will be paid to you by bank transfer -we will ask you to enter your bank details just before the session ends.

..... *Screen 1* .....

Please answer the following questions:

- How old are you? _____
- What is your gender?
  - o Female
  - o Male
  - o Other
- What is/was your field of study?
  - o Arts, Letters and Languages
  - o Law, Politics, Economics and Business
  - o Social and Human Sciences
  - o Techno, Health and Engineering Sciences
  - o Other _____

..... *Screen 2* .....

Choose a new first name that you will keep throughout the session.

- *For those who previously answered "Female":*
  Choices: Rose, Louise, Mia, Alice, Anne
- *For those who previously answered "Male":*
  Choices: Lucas, Raphaël, Adam, Louis, Arthur
- *For those who previously answered "Other":*
  Choices: Rose, Lucas, Louise, Raphaël, Mia, Adam, Alice, Louis, Anne, Arthur

..... *Screen 3* .....

Task instructions

In this part, at each stage, you will have to solve the sum of five randomly drawn two-digit numbers during 3 minutes. Only the paper provided is allowed. Please submit your answers by clicking on the "**Next**" button with your computer mouse, or by pressing the "**Enter**" key. No feedback about the correctness of your answers will be given. Stage 1 is about to start.

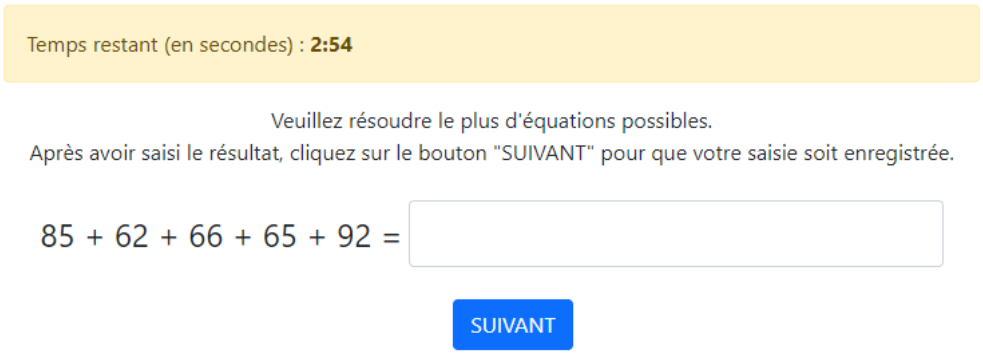
..... *Screen 4* .....

Stage 1: Piece-rate instructions

You will earn 0.50 € for each sum correctly solved during the 3'. Your payment does not decrease if you submit an incorrect answer. We will refer to this payment as the "**piece-rate payment scheme**". Your performance in this stage will be revealed to you at the end of Part 1, after completing all stages.

Example 1: If no answer is correct, your payment at this stage will be 0.00 €.
If you correctly solve 1 sum, your payment at this stage will be 0.50 €.
If you correctly 2 sums, your payment at this stage will be 1.00 €.
And so on and so forth.

Reminder: Only one of the four stages will be drawn to determine your final payment.



New screen

Stage 2: Tournament instructions

You perform the same kind of task during 3'. Now your payment will depend on your performance compared to the performance of a randomly chosen opponent. You will keep the **same opponent** until the end of the session. For this Stage 2, your earnings will depend on the number of sums you have correctly solved compared to your opponent. The one that has more correct answers receive 1.50 € per correct answer, and the opponent receives nothing. In the case of a tie, the winner will be randomly chosen. We refer to this payment as the "**tournament payment scheme**".

Example 2.1: If you have 4 correct answer and your opponent 3, you win 6.00 € and your opponent wins 0.00 € at Stage 2.
Example 2.2: If you have 3 correct answers and your opponent 6, your opponent wins 9.00 € and you win 0.00 € at Stage 2.

Your performance in this stage will be revealed to you at the end of Part 1, after completing all the stages. **Please note:** you will have to guess whether you won or lost in this stage. If your guess is correct, you win an extra 1.00 €.

Example 2.3: If you answer "Yes, I think I have won" and you actually have won, you

win 1.00 € in addition to your earnings for the game.

Example 2.4: If you answer "Yes, I think I have won" and you have actually lost, you win 0.00 €.

Reminder: Only one of the four stages will be drawn to determine your final payment.

..... *New screen* .....

Your opponent "***opponent's name***" has made a choice. You can now start the tournament.

..... *New screen* .....

You have to guess whether you lost or won. A correct guess will give you an extra 1.00 €. Do you think you have won the tournament at Stage 2?

- Yes
- No

..... *New screen* .....

Stage 3: Choice of payment method

At this stage, you have to perform the same task during 3'. However, at this stage, you have to choose the method in which you would like to be paid, either the one used in Stage 1 or that of Stage 2. Therefore, you have to choose between these two options:

Option 1: Piece-rate

If you choose this option, you will receive 0.50 € per correct answer in the task.

Option 2: Tournament

By choosing this option, your performance will be compared to the performance of your opponent in the previous Stage 2 you just performed. This ensures that, if you choose tournament,

you can play even if your opponent chooses the piece-rate. If you win the tournament you will earn 1.50 € for each sum correctly solved. If you choose this option and you do not win, your earnings in this stage will be zero. In the case of a tie, the winner will be randomly chosen. Again, also at this stage, you will have to guess whether you were the winner or the loser in this stage; a correct guess will ensure you an additional 1.00 €. Immediately after, you will be informed of your performance in this Stage 3.

Example 3.1: If you choose the piece-rate and solve 3 sums correctly, you earn 1.50 € at Stage 3.
Example 3.2: If you choose the tournament and solve 3 sums correctly while your opponent solved 6 sums correctly in Stage 2, you earn 0.00 € at Stage 3.
Example 3.3: If you choose the tournament and solve 6 sums correctly while your opponent has solved 3 in Stage 2, you earn 9.00 € at Stage 3.

..... *New screen* .....

Stage 4: specific instructions [Stage 4.a: Switching point instructions]
At this stage, you will be asked whether you "accept" or "not accept" to compete again against the same opponent for various point gaps. We will show you several examples of the different ways of determining your earnings.

You will then have 3 minutes to perform the task of two-digit numbers sum.

Reminder: Only one of the four stages will be drawn to determine your final payment.

..... *New screen* .....

For winners:

Would you "accept" or "not accept" to compete again against **Louis**, according to the following scenaria from the previous stage?

	I accept	I do not accept
You have solved correctly the same number of sums, so you are tied	◉	○
You have solved correctly 1 more sum	◉	○
You have solved correctly 2 more sums	◉	○
You have solved correctly 3 more sums	◉	○
You have solved correctly 4 more sums	◉	○
You have solved correctly 5 more sums	◉	○
You have solved correctly 6 more sums	◉	○
You have solved correctly 7 more sums	◉	○
You have solved correctly 8 more sums	◉	○
You have solved correctly 9 more sums	◉	○
You have solved correctly 10 more sums	◉	○

For losers:

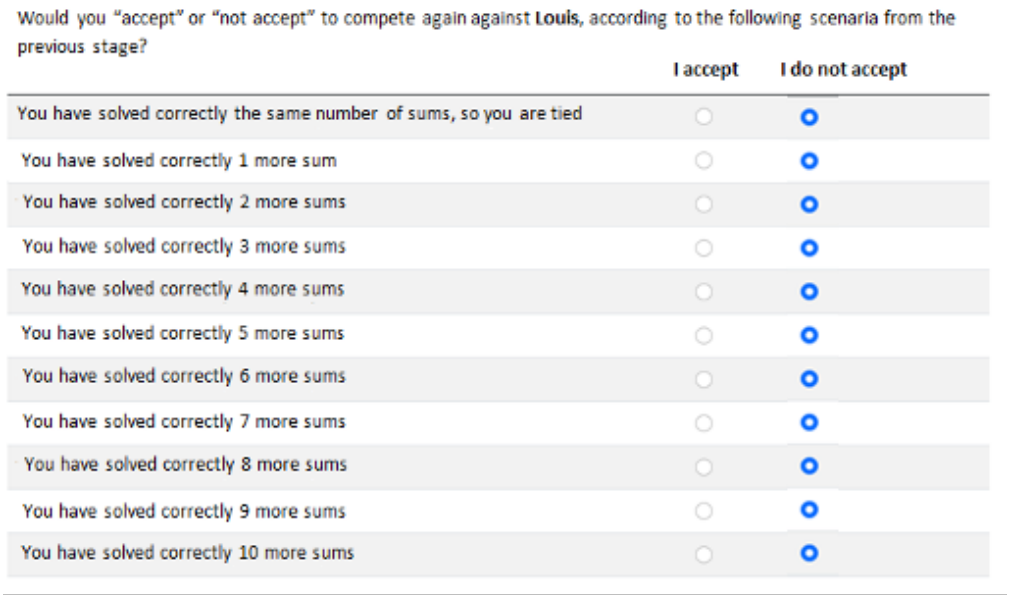
Would you "accept" or "not accept" to compete again against **Louis**, according to the following scenaria from the previous stage?

	I accept	I do not accept
You have solved correctly the same number of sums, so you are tied	◉	○
Your opponent solved correctly 1 more sum	◉	○
Your opponent solved correctly 2 more sums	◉	○
Your opponent solved correctly 3 more sums	◉	○
Your opponent solved correctly 4 more sums	◉	○
Your opponent solved correctly 5 more sums	◉	○
Your opponent solved correctly 6 more sums	◉	○
Your opponent solved correctly 7 more sums	◉	○
Your opponent solved correctly 8 more sums	◉	○
Your opponent solved correctly 9 more sums	◉	○
Your opponent solved correctly 10 more sums	◉	○

Example 4.1: If you choose "I accept" for every possible point gaps your performance in this Stage 4 will be compared to the performance of your opponent in Stage 2. If you win the tournament (i.e. you perform more correct sums than your opponent in Stage 2), you will receive 1.50 € per correct sum. If you do not win, you will receive nothing at this Stage 4. In the case of a tie, the winner will be randomly chosen. Also, you will have to guess whether you won or lost in this stage. If your guess is correct, you will win an extra 1.00 €. Immediately after, you will be informed of your performance in this tournament.

. . . . . *New screen* . . . . . .

For winners:



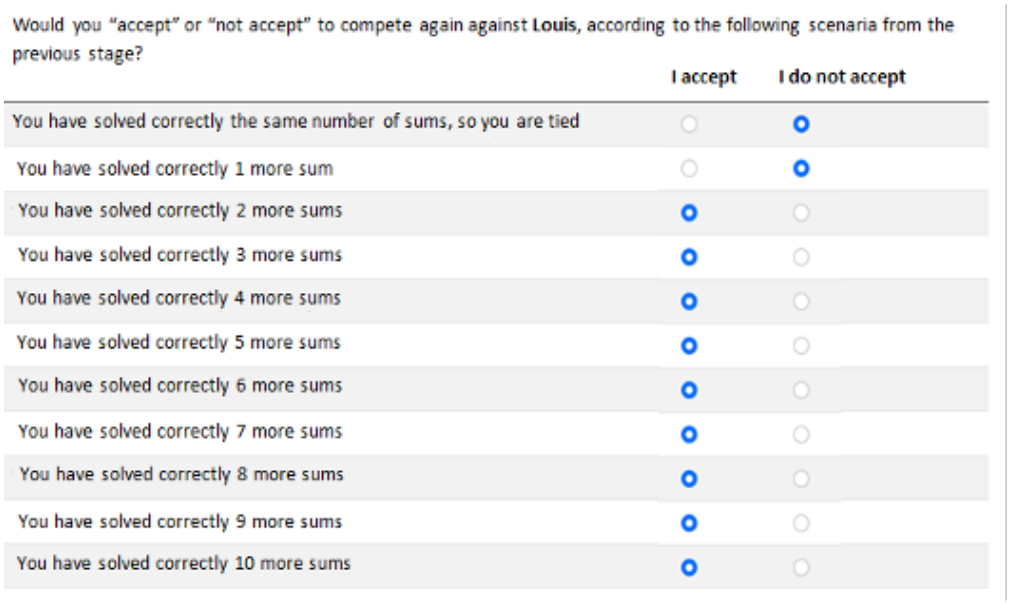
For losers:



Example 4.2: If your choice is "I do not accept" for any possible point gap, you will be paid according to the piece-rate payment scheme in this Stage 4. You will receive 0.50 € per correct answer.

New screen

For winners:



Example 4.3: If you choose "I do not accept" for a 1-point gap and you have actually solved correctly 1 more sum than your opponent in Stage 3, you will be paid according to a piece-rate payment scheme in Stage 4.

Example 4.4: If you choose "I accept" for a 2-points gap and you have actually solved correctly 2 more sums than your opponent in Stage 3, you will be paid according to a tournament payment scheme in Stage 4.

For losers:



Would you "accept" or "not accept" to compete again against **Louis**, according to the following scenaria from the previous stage?

	I accept	I do not accept
You have solved correctly the same number of sums, so you are tied	●	○
Your opponent solved correctly 1 more sum	●	○
Your opponent solved correctly 2 more sums	●	○
Your opponent solved correctly 3 more sums	○	●
Your opponent solved correctly 4 more sums	○	●
Your opponent solved correctly 5 more sums	○	●
Your opponent solved correctly 6 more sums	○	●
Your opponent solved correctly 7 more sums	○	●
Your opponent solved correctly 8 more sums	○	●
Your opponent solved correctly 9 more sums	○	●
Your opponent solved correctly 10 more sums	○	●

Example 4.5: If you choose "I do not accept" for a 3-points gap and your opponent has actually solved correctly 3 more sums than you in Stage 3, you will be paid according to a piece-rate payment scheme in Stage 4.

Example 4.6: If you choose "I accept" for a 2-points gap and your opponent has actually solved correctly 2 more sums than you in Stage 3, you will be paid according to a tournament payment scheme in Stage 4.

. . . . . *New screen* . . . . . .

Decision screen for winners :

Would you "accept" or "not accept" to compete again against **Louis**, according to the following scenaria from the previous stage?

	I accept	I do not accept
You have solved correctly the same number of sums, so you are tied	○	○
You have solved correctly 1 more sum	○	○
You have solved correctly 2 more sums	○	○
You have solved correctly 3 more sums	○	○
You have solved correctly 4 more sums	○	○
You have solved correctly 5 more sums	○	○
You have solved correctly 6 more sums	○	○
You have solved correctly 7 more sums	○	○
You have solved correctly 8 more sums	○	○
You have solved correctly 9 more sums	○	○
You have solved correctly 10 more sums	○	○

Decision screen for losers :

Would you "accept" or "not accept" to compete again against **Louis**, according to the following scenaria from the previous stage?

	I accept	I do not accept
You have solved correctly the same number of sums, so you are tied	○	○
Your opponent solved correctly 1 more sum	○	○
Your opponent solved correctly 2 more sums	○	○
Your opponent solved correctly 3 more sums	○	○
Your opponent solved correctly 4 more sums	○	○
Your opponent solved correctly 5 more sums	○	○
Your opponent solved correctly 6 more sums	○	○
Your opponent solved correctly 7 more sums	○	○
Your opponent solved correctly 8 more sums	○	○
Your opponent solved correctly 9 more sums	○	○
Your opponent solved correctly 10 more sums	○	○

[**For those who chose "I accept" for the actual point gap**

..... *New screen* .....

In Stage 3, the points gap with your opponent was x point(s). For this gap, your answer was: "I accept". Therefore, at Stage 4 you will perform the task under the tournament payment.

As a reminder, you have 3 minutes to perform the task. Your earnings will depend on the number of correct sums you have solved compared to your opponent at Stage 2. The participant who solves correctly more sums will win the tournament and receive 1.50 € per correct answer, while the loser receives nothing. In the case of a tie, the winner will be randomly chosen. Also, you will have to guess whether you won or lost in this stage. If your guess is correct, you will win an extra 1.00 €. Immediately after you will be informed of your performance in the tournament.

Reminder: Only one of the four stages will be drawn to determine your final payment.

..... *New screen* .....

Your opponent "***opponent's name***" has made a choice.

..... *New screen* .....

You have to guess whether you lost or won. A correct guess will give you an extra 1.00 €. Do you think you have won the tournament at Stage 2?

- Yes
- No

..... *New screen* .....

"*You have won*" or "*You have lost*"]

[For those who chose "I do not accept" for the actual point gap

..... *New screen* .....

In Stage 3, the points gap with your opponent was x point(s). For this gap, your answer was: "I do not accept" to compete." Therefore, you will perform the task under the piece-rate payment.

As a reminder, you have 3 minutes to perform the task. You earn 0.50 € per correct answer.

Reminder: Only one of the four stages will be drawn to determine your final payment.]

[For the ones that chose option 1 in Stage 3]

..... *New screen* .....

You have chosen the piece-rate. As a reminder, you have 3 minutes to perform the task. You will get 0.50 € per correct sum.

Reminder: Only one of the four stages will be drawn to determine your final payment.

..... *New screen* .....

Stage 4: Specific instructions [Stage 4b WTP instructions]

At Stage 3 you chose the piece-rate payment scheme. However, you will be asked to perform the same task under a tournament payment scheme, having the possibility to avoid the competition by giving up an amount of your total gains. The next screen will ask you to choose

the **maximum** amount you are willing to pay to avoid the tournament from the interval [0.25 €, 2 €] in steps of 0.25. Simultaneously, an amount of money within the same interval [0.25 €, 2 €] will be randomly drawn. That amount will determine your real payment scheme at Stage 4, as follows: two outcomes are possible:

Outcome 1

If the amount you have chosen is smaller than the amount randomly drawn, you will not avoid to compete and, therefore, be paid according to the tournament payment scheme. Your performance will be compared with your opponent's performance, "***opponent's name***", in Stage 2. If your correct answers are higher than your opponent at Stage 2, you will receive 1.50 € per correct answer. You will get nothing otherwise. In the case of a tie, the winner will be randomly chosen. Also, you will have to guess whether you won or lost in this stage. If your guess is correct, you will win an extra 1.00 €. Immediately after, you will be informed of your performance in the tournament at Stage 4.

Example 4.7: You choose 0.25 € and the amount drawn is 2.00 €. Therefore, you enter the tournament at Stage 4. At this stage, you solve correctly 6 sums while your opponent correctly solved 10 in Stage 2. You get 0.00 € in Stage 4. If, on the contrary, your correct answers were 11 and those of your opponent are 10, then you win the tournament and get 11x1.50 = 16.50 €.

Outcome 2

If the amount you have chosen is higher or equal than the amount randomly drawn, you will avoid the competition and, therefore, you will be paid according to the piece-rate payment scheme. You will earn 0.50 € per correct answer.

Example 4.8: You choose 1.25 € and the amount drawn is 0.50 €. If you have 10 correct answers, you get 0.50x10= 5.00 €.

In both cases, the amount drawn will be deducted from your earnings for that part of the experiment. Once you have made your choice, you will have to repeat the same task as before. You will have 3 minutes to perform the task.

Reminder: Only one of the four stages will be drawn to determine your final payment.

..... *New screen* .....

What is the maximum amount (of your earnings) that you are willing to give up to avoid competition?

- 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 2.00

[For those who chose a WTP $<$ WTP random

..... *New screen* .....

The amount randomly drawn is Y €. The maximum amount you are willing to pay to avoid competition is X €.
The amount you have chosen is lower than Y. You have to play under a tournament payment scheme at Stage 4. The conditions are the same as in previous stages with tournament.

..... *New screen* .....

Your opponent "***opponent's name***" has made a choice.

..... *New screen* .....

You have to guess whether you lost or won. A correct guess will give you an extra 1.00 €. Do you think you have won the tournament?

- Yes

- No

..... *New screen* .....

"*You have won*" or "*You have lost*"]

[For those who chose a WTP $\geq$ WTP random

..... *New screen* .....

The amount randomly drawn was Y €. The maximum amount you are willing to pay to avoid competition is x €.

The amount you have chosen is higher than Y. You will play under a piece-rate payment scheme in Stage 4. The conditions are the same as in previous stages with piece-rate payment.]

Instructions for Part 2

[Subjects perform first the Bomb Risk Elicitation Task (BRET)]

..... *New screen* .....

In the last part of the experiment, you will have to answer a few questions about the presumed subject of our study and the games you have just played. You will then be asked to answer a questionnaire.

..... *New screen* .....

What do you think this study is about?

..... *New screen* .....

Why did you (or did not) choose the tournament in Stage 3?

_____

What was the gender of your opponent?

- Female
- Male
- I do not know

What was the name of your opponent?

_____

What was your final payoff?

_____

..... *New screen* .....

[Subjects now perform the Gender Role Attitudes Scale (GRAS)]

Chapter 2

Perceptions of inequality: When the grass is less green on the other side

"*Injustice anywhere is a threat to justice everywhere.*" – Dr. Martin Luther King, JR. (Letter from Birmingham Jail, 1963)

2.1 Introduction

Despite the predictions of the Kuznets (1955) curve suggesting that inequality diminishes as income rises, developed countries have recently experienced a resurgence in inequality. Chancel *et al.* (2022) argue that while inequality between countries has declined, the level of inequality within countries has reached unprecedented heights. Beyond social justice purposes (Rawls, 1971), inequality is detrimental in several aspects: health, social cohesion, and general well-being (Wilkinson & Pickett, 2017; Ugur, 2021). Moreover, inequality leads to increased crime and negatively affects economic growth (Benabou, 1996; Knell, 1999; Kelly, 2000; Stiglitz, 2015). According to Chancel *et al.* (2022), "*an accelerated compression of inequality among and within countries will require a massive redistribution of wealth*" (p.69). However, the perception of inequality influences the desire for redistribution (Hauser & Norton, 2017) and recent studies report a tendency to underestimate wealth inequality in developed countries (Osberg & Smeeding, 2006; Norton & Ariely, 2011; Niehues, 2014; Gimpelson & Treisman, 2018; Bavetta *et al.*, 2019). For instance, according to a survey conducted in France, Germany, and Italy, 43% of respondents consider that gender equality has been achieved[1,2]– although women still earn on average 12.70% less per hour than men in the European Union (European Commission, 2021). In the United States, individuals tend to overestimate the progress of Black-White economic equality (Kraus *et al.*, 2017).

Misconceptions could arise from the fact that economic growth is associated with a reduction in inequality (Kuznets, 1955), which can prevent rich regions from acknowledging the persistence or even increase of inequality. Misconceptions could also be induced by temporal comparisons, as evidenced by Spoor & Schmitt (2011). The authors demonstrate that recalling past discrimination tends to minimize current discrimination, especially for the victim group – in their case, women. This psychological mechanism of invoking past experiences to

[1]See https://focus2030.org/Sondage-Inegalites-femmes-hommes-constitution-IVG-diplomatie-feministe-quelles

[2]The question asked was "In your opinion, has equality between men and women been achieved?" without specifying equality of rights or income.

promote the present is commonly known as the "things were *worse* before" approach.[3] The present study experimentally addresses another form of comparison that fosters misconceptions of the level of inequality: social comparison with others, or the belief that "things are *worse* somewhere else". For instance, what would be the beliefs regarding gender equality in Europe if the situation of European women were compared to that of Afghan women?

The concept of inequality covers many dimensions – inequality of rights, opportunity, or income. I choose to focus on wealth inequality between individuals, without considering the gender or ethnic dimensions, for instance. The aim of this chapter is to explore how the level of wealth inequality observed in other groups influences the willingness of wealthier individuals to reduce inequality within their own group. Therefore, the first research question can be expressed as follows. Do wealthier individuals redistribute less when they observe greater inequality in another group? For this purpose, I study redistribution from the wealthiest to the poorest player within a group using a modified Dictator Game in a laboratory experiment. In one of the treatments (Info Inequality), subjects are informed of the higher level of inequality in another group. In a second treatment (Info Actions), I disclose the action of a more unequal group not to redistribute. This aims to investigate the second research question: Do wealthy individuals become less inclined to redistribute when they are informed that the wealthier in a more unequal group choose not to redistribute? In order to control for the reference point chosen by the rich in their redistribution choice, two additional treatments respectively disclose only the allocation of the richest (Info Rich) and poorest (Info Poor) members of the most unequal group. Experiments benefit from a controlled environment to provide different levels of inequality and types of information to determine whether people act to restore justice. The neutral framework allows for isolating the factors that determine perceptions of inequality. This is particularly relevant in this context since several factors are determinants of the subjective vision of inequality (Bavetta *et al.*, 2019). Through the

[3]In contrast, there is also a "things were *better* before" approach (Serres, 2017). According to psychology, this pattern can be explained by a negativity bias: humans tend to give greater weight to negative events (Rozin & Royzman, 2001). However, this sensitivity to negative events is instantaneous, whereas our memory tends to retain positive events and forget negative ones (e.g., Denny & Hunt, 1992; Sedikides & Green, 2009). This effect is also known as the "rosy retrospection" (Mitchell *et al.*, 1997).

different treatments, I examine the poorest player's perception of fairness regarding the allocation received in order to assess the third research question. Do poorer individuals tend to perceive inequality within their group as fairer when they observe a *worse* situation in another group, either in terms of initial allocation or decision not to redistribute?

This chapter provides evidence that the disclosure of others' inequality negatively impacts the actions to reduce inequality within the group. Data shows that rich tend to choose a less equal distribution when aware of larger inequality in another group. This demonstrates that social comparison influences the willingness to (not) move towards equality. Consequently, this raises questions about the efforts deployed to reduce inequality in a globalized world, where comparisons between countries are facilitated by the media, for example. However, actions to restore in-group equality seem unrelated to those perceived in other groups. The decision-maker does not redistribute less when observing that another more unequal group chooses not to. On the other hand, recipients perceive the allocation received as fairer when the rich's allocation or the decision not to redistribute in the other group are disclosed.

The rest of the chapter is structured as follows. Section 2.2 provides a review of relevant literature. Section 2.3 presents the design and Section 2.4 describes the hypotheses. The results are presented in Section 2.5 and Section 2.6 discusses the potential mechanisms. Section 2.7 concludes the chapter.

2.2 Literature

Perceptions of inequality are mostly studied through the scope of fairness views (e.g., Almås *et al.*, 2010; Bavetta *et al.*, 2019; Ugur, 2021; Du & King, 2022). Citizens differ in their perceptions or tolerance of inequality since they hold different views on what constitutes *fair* inequality (Almås *et al.*, 2010).[4] Using the International Social Survey Programme's data,

[4]In their article, Almås *et al.* (2010) discuss what could be considered as *fair* inequality, such as those due to chance or differences in productivity. They study children's redistribution choice in a dictator game with a production phase, allowing them to distinguish between egalitarianism (considering all inequalities as unfair),

Bavetta *et al.* (2019) identify two patterns: Nordic countries, which tend to underestimate inequality, and Mediterranean countries, which have higher perceptions of inequality. Their results demonstrate that a higher level of perceived inequality is associated with greater perceived unfairness, except in Great Britain. The British tend to perceive inequality as high but not unfair, mainly because they believe that their country offers equal opportunities to all. By shaping perceptions of inequality, conceptions of fairness also affect tolerance. For instance, Alesina *et al.* (2004) demonstrate that the poor in Europe are more concerned about inequality than the poor in the United States. This higher tolerance among the US citizens lies in their beliefs about mobility within their society; even though they tend to overestimate it (Davidai & Gilovich, 2015). The underestimation of inequality or the overestimation of mobility is a threat to equality since it impacts support for redistribution within the population (Alesina & Angeletos, 2005; Krawczyk, 2010; Mollerstrom *et al.*, 2015; Müller & Renes, 2021). Biased beliefs regarding mobility can even lead to positions that contradict the individual's status. The "Prospect of upward mobility" (POUM) states that optimistic beliefs about the possibility of moving up the income distribution lead to lower support for redistributive policies, such as taxation (Putterman, 1997; Benabou & Ok, 2001). These visions also determine the decision to redistribute resources, as does the social system adopted by countries: egalitarianism, meritocratic, or libertarianism. Consequently, the adoption of a particular system shapes the perceptions of inequality and they could change with a transformation in governance. For example, Kelley & Zagorski (2004) demonstrate that Central-East European egalitarian norms changed after the collapse of Communism, leading to a greater acceptance of inequality. Other factors have been found to play a role in the perception of inequality, such as age (Arsenio & Willems, 2017), political orientation (Norton & Ariely, 2011; Kraus *et al.*, 2017; Waldfogel *et al.*, 2021) or position on the income scale (Knell & Stix, 2020).

Numerous survey studies demonstrate that lower-ranked respondents tend to overestimate their position, while higher-ranked respondents underestimate their own (Cruces *et al.*, 2013;

meritocratism (justifying those resulting from differences in production), and libertarianism (justifying all inequalities in earnings).

Karadja *et al.*, 2017; Gimpelson & Treisman, 2018; Hvidberg *et al.*, 2020; Knell & Stix, 2020). This bias is driven by the desire to improve one's own perceived position, by underestimating others' positions (Hvidberg *et al.*, 2020). Social comparison, and especially being ranked above peers, increases happiness (Clark & Oswald, 1996; Ferrer-i-Carbonell, 2005; Clark *et al.*, 2008; Clark & Senik, 2010). Indeed, individuals appear more sensitive to relative income compared to absolute income (e.g., Easterlin, 1974; Alderson & Katz-Gerro, 2016). The theory of *relative* income constitutes one of the plausible answers to Easterlin's paradox. Economic growth does not necessarily increase happiness if the individual's income increases at the same rate as others (Easterlin, 1974). Apart from well-being, surveys demonstrate that the subjective position of citizens in the income distribution explains their demand for redistribution (Clark & Senik, 2010; Cruces *et al.*, 2013; Karadja *et al.*, 2017; Hoy & Mager, 2021). This outcome is theoretically supported by Meltzer & Richard (1981): the median voter is favorable to the redistribution if their income is below the mean. Regarding the perceptions of inequality between countries, citizens may overestimate others' inequality (or underestimate theirs) to appear superior, which may prevent efforts to move towards equality. Therefore, investigating the role of social comparison in the perception of inequality appears crucial to achieving equality. Due to the multidimensionality of inequality, measuring perceptions is complex (Bavetta *et al.*, 2019). The use of lab experiments, which offer a neutral framework in which we can control for several aspects, appears pertinent.

Social comparison is addressed in experimental studies of giving and sharing behaviors, using Ultimatum Games (Güth *et al.*, 1982) or Dictator Games (Forsythe *et al.*, 1994), henceforth UG and DG respectively. Previous research already provided insight into the decision to donate during these games. Following the theory, the best strategy is to give nothing (in the DG) or the smallest amount (in the UG). However, equal distribution is often preferred (for a review, see Camerer, 2003). The motives commonly studied are generally altruism (Andreoni & Miller, 2002), inequality aversion (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000), or social norms (e.g., Bohnet & Zeckhauser, 2004; Bicchieri & Xiao, 2009;

Binmore, 2010).[5] Social comparison is advanced to explain a higher rejection rate in UG (Knez & Camerer, 1995) or higher donation in DG (Cason & Mui, 1998).[6] In psychology, the expectation of a *fair* offer induces a significantly higher rejection rate in UG (Sanfey, 2009). For this chapter, I choose to use a DG to study the willingness to reduce inequality through redistribution. The choice of this game is motivated by the fact that donations in the lab during a DG can reflect preferences for redistribution in the field (Tammi, 2013). Furthermore, this rules out the dictator choosing the distribution in anticipation of potential rejection by the recipient, as in the UGs. This research contributes to the literature of Almas *et al.* (2021, 2022) e.g., that assess redistribution using a DG with a spectator. Like their *Luck* treatments, I randomly allocate unequal endowments to recipients. I believe that this procedure reproduces the random nature of birth, i.e. being born in a more or less unequal country or into a more or less wealthy family. Moreover, I create the level of inequality rather than making it dependent on the decisions of dictators. However, as opposed to Almas *et al.* (2021, 2022), the decision to redistribute relies on the player experiencing the advantageous inequality and not on a third party. This likely results in the decision to depend on a status quo bias (Samuelson & Zeckhauser, 1988) or a desire for relative superiority compared to the other group. I control for the latter by assessing the reference point chosen for comparison, as it matters for well-being and redistributive preferences (e.g., Kuziemko *et al.*, 2014; Alderson & Katz-Gerro, 2016; Karadja *et al.*, 2017). In the following section, I describe the experimental design before presenting the hypotheses.

[5]For a review of offers' determinants, see the meta-analysis of Engel (2011).

[6]Note that I distinguish between social comparison and social norm. In Bohnet & Zeckhauser (2004) and Sanfey (2009), the social norm is induced by the disclosure of the *average* offer made. On the other hand, Knez & Camerer (1995) and Cason & Mui (1998) disclose the offer made by *one* other proposer.

2.3 Experimental Design

2.3.1 The task

The experiment is based on a between-subjects design. Subjects are divided into groups of two players and perform a modified DG (Forsythe *et al.*, 1994). Roles are randomly assigned between Player 1 (the wealthiest) and Player 2 (the poorest). Table 2.1 presents the options of distribution. Subjects receive initial allocations (Option A) implying inequalities within their group. In the next screen, the richest player (Player 1) is asked to decide if they want to keep allocations unchanged (Option A) or if they want to redistribute equitably the resources (Option B), and it is framed as so.

In Group 1, Player 1 first receives 6 € and their partner 3 € (Option A). Player 1 has the choice to redistribute the joint payoff of the group so that both players receive 4.50 € each (Option B). Simultaneously, in another group (Group 2), the wealthiest player is asked to make a similar choice but with different allocations. In Group 2, Player 1 has the choice to keep 6 € for themself and let 1 € to the other player (Option A) or to receive 3.50 € each (Option B). In each group, the initial allocation of Player 2 is inferior to that of Player 1. In Group 2, the difference between the two players' allocations is 80% while, in Group 1, Player 1 receives only 50% more than Player 2. Therefore, Group 2 is initially more unequal than Group 1. Depending on the treatment detailed below, Group 1 observes the allocations or the decision made in Group 2. Note that Group 2 could be considered a passive group as it mainly serves to disclose information to Group 1.

Table 2.1: Options of distributions

		Group 1	Group 2
Option A	Player 1	6 €	6 €
	Player 2	3 €	1 €
Option B	Player 1	4.50 €	3.50 €
	Player 2	4.50 €	3.50 €

2.3.2 Treatments: Disclosure of others' inequality

Among all conditions, Player 1 and Player 2 know both the allocations and the decision (not) to redistribute within their own group. In the treatments, I disclose different levels of information about Group 2 to Group 1.[7]

- **Baseline**: Group 1 does not have any information about Group 2.
- **Info Poor**: the initial allocation (Option A) of the poorest in Group 2 (i.e. 1 €) is disclosed to Group 1. Subjects are not aware whether it is the allocation of the poorest and they are asked to guess the status of the player (i.e. poorest or richest) whose allocation is disclosed (henceforth Status Guess). This treatment induces a reference point for players, a comparison with *worse* payoff.
- **Info Rich**: the initial allocation (Option A) of the richest in Group 2 (i.e. 6 €) is disclosed to Group 1. As in Info Poor, subjects have to guess the status of the player (Status Guess) since I only disclose the allocation without specifying if it is the richest or the poorest player in Group 2. As a complement to Info Poor, this treatment allows the comparison for Player 1 with similar payoff and with *better* payoff for Player 2.
- **Info Inequality**: the initial allocations (Option A) of both Players 1 and 2 from Group 2 (i.e. 6 € and 1 €) are disclosed. This treatment enables intergroup inequality comparisons since, as opposed to Info Poor and Info Rich, subjects of Group 1 are aware of the payoff difference between Players 1 and 2 in Group 2. More specifically, it induces comparison with a group experiencing a more pronounced level of inequality between the two players.
- **Info Actions**: Participants in Group 1 have first to guess whether Player 1 in Group 2 is going to choose Option A or Option B (henceforth Choice Guess). Then, subjects in

[7]During the sessions, Group 2 also receives the same type of information about Group 1 (except for Info Actions). Due to budget constraints and to maximize the statistical power of the analysis, I choose to have only one Group 2 in each session. Since one session was conducted per condition, there are 5 Group 2s in the sample. Therefore, the analysis focuses on the behavior of Group 1 after receiving information about Group 2.

Group 1 are informed that Group 2 decided not to redistribute. Note that one Group 2 was actually present for this treatment (as for the others) and the decision not to redistribute is the actual decision made by the Group 2 present in the session.

Across all conditions, after receiving information about Group 2 and the allocation corresponding to the option chosen by Player 1, Player 2 is asked whether they perceive the allocation received from Player 1 as fair or unfair (henceforth Fairness Perception), and why.[8]

2.3.3 Questionnaire

At the end of the experiment, subjects have to report sociodemographic characteristics (age, gender, and education). I control for subjects' perceptions of inequality between countries with questions from the International Social Survey Programme's 2019 "Social Inequality" (ISSP) (e.g., Roberts *et al.*, 2023) and within the society, using questions from Kahan *et al.* (2011). The first set of questions allows capturing whether subjects believe the economic differences between rich and poor countries are too large (henceforth Intolerance Global Inequalities) or if the differences between poor and rich countries are fair (henceforth Fairness belief of Global Differences). It also allows knowing if the subjects are in favor of a tax from rich countries to poor countries (henceforth Global Tax Support) or of migration from poor countries (henceforth Migration support). The survey from Kahan *et al.* (2011) enables to classify subjects between egalitarian or hierarchical worldview.[9] Moreover, subjects have to complete the Social Value Orientation Test (henceforth SVO) from Murphy *et al.* (2011). This test allows classifying subjects into the following categories: competitive, individualistic, altruistic, and prosocial. For an overview of the instructions, please refer to Appendix 2.D.

[8]Note that, as for Player 1, Player 2 first saw a screen with the initial allocation (Option A) and is aware that Player 1 could change the distribution. At the end of the DG's part, Player 2 know the both options and the option chosen.

[9]According to Kahan *et al.* (2011), hierarchical and egalitarian worldview refer to "*attitudes toward social orderings that connect authority to stratified social roles based on highly conspicuous and largely fixed characteristics such as gender, race, and class*" (p.206). While hierarchicals are in favor of distinctions (of wealth, for example), egalitarians believe that distinctions between individuals should be minimized.

2.3.4 Procedure

The experiment was pre-registered at AEA RCT Registry, and approved by the Ethics Committee of the University Jaume I of Castellón (Spain).[10] The sessions were conducted at the Laboratorio de Economía Experimental (LEE) at the University Jaume I of Castellón (Spain) during March 2024. The experiment was computerized using the oTree software (Chen *et al.*, 2016). The experimental sessions lasted approximately 15 minutes and subjects were paid on average 7.47 €. Participants' final payoff corresponded to the amount received in the task and the show-up fee of 3.00 €. The sample includes 300 participants, with 59.33% identifying as female, 39.33% as male and 1.33% not reporting their gender.

2.4 Hypotheses

This study focuses on actions taken to reduce inequality, specifically the willingness to change the status quo for a more equitable distribution within the group. Since sticking with the original allocation is costless (Samuelson & Zeckhauser, 1988), redistribution choices are mainly driven by inequality aversion (Fehr & Schmidt, 1999). This section investigates whether the disclosure of different information regarding Group 2 induces a behavioral shift among players in Group 1. The hypotheses are derived from the review of literature and are tested in the subsequent analysis.

When a decision-maker in a group is aware of greater disparities in another group (Info Inequality), they are expected to choose equal distribution less frequently than in the absence of this information. A study of the UG from Sanfey (2009) suggests that individuals are more inclined to accept unfair offers if they anticipate receiving low offers. In the present chapter, the comparison with a higher level of inequality is expected to reduce the dictator's willingness to move towards equality. Additionally, Karadja *et al.* (2017) indicates that the

[10]Registry number AEARCTR-0013113 (link: https://www.socialscienceregistry.org/trials/13113)

demand for publicly administered redistribution diminishes when individuals are informed of their higher rank relative to others, while those with lower ranks tend to advocate for more redistribution (Cruces *et al.*, 2013). The richest player in both Groups 1 and 2 should be less in favor of the redistribution compared to the poorest. Nevertheless, the decision-maker in Group 1 perceives that their group requires less redistribution compared to Group 2 because the intergroup comparison in Info Inequality reshapes their perceptions of inequality, consequently influencing their tolerance towards them.

Hypothesis 1: *The richest player, i.e. Player 1, in Group 1 redistributes less often when aware of Group 2's more pronounced inequality (Info Inequality).*

In Info Actions, subjects are informed that the dictator in Group 2 did not redistribute. The richest player in Group 1 is then predicted to redistribute less because they benefit from *moral wiggle room* not to redistribute (Dana *et al.*, 2007). In Dana *et al.* (2007)'s experiment, the lack of transparency (regarding the outcome) increases selfishness while maintaining the illusion of fairness. The paper concludes that "*people feeling compelled to give due to situational factors, while not really valuing the corresponding outcome*" (p.77). When Group 2 does not move towards equality, it allows the rich in Group 1 not to opt for the equitable outcome. According to Bandura & Walters (1963), observation of others allows to play a strategy that the player has not dared to play. If a player observes that another player receives a higher payoff when deviating from the subgame perfect strategy in a UG, they are more likely to adopt the same strategy (Duffy & Feltovich, 1999). In this game, observing others' decision not to redistribute results in the equitable situation being less chosen by the rich in Group 1. Moreover, the decision-maker might choose not to redistribute to preserve their higher status (Ferrer-i-Carbonell, 2005; Clark & d'Ambrosio, 2015). As presented in Section 2.3, if the decision-maker chooses the equitable situation, their allocation is inferior to that of the decision-maker in the other group. To maintain their well-being, the richest player in Group 1 should not redistribute when Group 2 chooses not to redistribute (Clark & Oswald, 1996; Ferrer-i-Carbonell, 2005; Clark *et al.*, 2008; Clark & Senik, 2010).

Hypothesis 2: *The richest player, i.e. Player 1, in Group 1 opts less often for the equal*

distribution when Group 2 does not redistribute (Info Actions).

In addition to examining the decision to redistribute resources equitably within a group, I investigate Player 2's perception of fairness. When the subject is only aware of the allocations received within their group, they tend to perceive the unequal allocation received as unfair. However, disclosing that another group experiences more pronounced inequalities is expected to increase their perception of fairness because they observe *worse* elsewhere. In Spoor & Schmitt (2011), the disadvantaged group tend to minimize the level of discrimination when the past (more pronounced) level of inequality is put in evidence. In this experiment, social comparison prompts recipients to evaluate their situation by referencing another group's inequality. An increase in fairness perception is also predicted when subjects are informed that the other group chooses not to redistribute. In the same manner as in Sanfey (2009), the recipients are informed that the unfair outcome was preferred in another group and they may expect to also receive the unequal distribution. This alignment of expectations with results reduces disappointment (Gill & Prowse, 2012), leading to a fairer perception regarding the allocation received.

Hypothesis 3: *Player 2s (in Group 1) perceive their own allocation as fairer when informed of the level of inequality (Info Inequality) or decision not to redistribute (Info Actions) in another group.*

2.5 Results

Table 2.2 provides a summary of descriptive statistics (Age, Female and Education) by treatment.[11] It also displays the proportion of Fairness Perception in Group 1 (takes the value 1 if Player 2 considers the allocation received as fair and 0 otherwise) and the proportion of

[11]Note that the overall sample is not gender-balanced, but if we consider the proportion of women assigned to Player 1, the sample is balanced (χ^2, $p = 0.51$). Subjects are on average older in Info Rich and Info Actions but the age is not correlated to the decision to redistribute equally, i.e. Option B (Spearman's $\rho = -0.08$, $p = 0.19$).

prosocial individuals, according to the SVO task.

Table 2.2: Summary of descriptive statistics

	Treatments						
	Baseline	Info Poor	Info Rich	Info Inequality	Info Actions	*p*	Total
Age	22.00 (4.79)	22.23 (5.43)	27.25 (12.72)	21.87 (5.52)	25.47 (10.96)	*0.00*	23.76 (8.77)
Female (%)	70.00 (0.46)	71.67 (0.45)	41.67 (0.50)	58.33 (0.49)	55.00 (0.50)	*0.03*	59.33 (0.49)
Fairness belief in Group 1 (%)	44.83 (0.51)	34.48 (0.48)	72.41 (0.45)	27.59 (0.45)	58.62 (0.50)	*0.00*	47.59 (0.50)
Prosocial (%)	73.33 (0.45)	71.67 (0.45)	71.67 (0.45)	60.00 (0.50)	56.66 (0.50)	*0.17*	66.67 (0.47)
Egalitarian (%)	50.42 (0.50)	50.17 (0.50)	49.72 (0.50)	50.39 (0.50)	49.71 (0.50)	*0.72*	49.92 (0.50)
Education (%)							
- Accounting and Finance	8.33	16.67	10.00	3.33	1.67	*0.02*	8.00
- Administration, Business and Management	21.67	16.67	6.67	21.67	8.33	*0.05*	15.00
- Arts, Languages and Communication	10.00	-	6.67	6.67	10.00	*0.17*	6.67
- Economics	10.00	6.67	6.67	11.67	15.00	*0.50*	10.00
- Education	6.67	5.00	5.00	3.33	8.33	*0.80*	5.67
- Health sciences	10.00	20.00	11.67	10.00	8.33	*0.30*	12.00
- Labour Relations and Human Resources	1.67	1.67	8.33	1.67	3.33	*0.18*	3.33
- Law, Criminology and Security	6.67	10.00	6.67	5.00	11.67	*0.65*	8.00
- Psychology	10.00	10.00	6.67	5.00	8.33	*0.82*	8.00
- STEM	8.33	10.00	15.00	28.33	20.00	*0.02*	16.33
- Other	6.67	3.33	16.67	3.33	5.00	*0.22*	7.00
Player 1s in Group 1	29	29	29	29	29		145
Player 2s in Group 1	29	29	29	29	29		145
Player 1 in Group 2	1	1	1	1	1		5
Player 2 in Group 2	1	1	1	1	1		5
Number of observations	60	60	60	60	60		300

Notes: Standard deviations are in parentheses. The p-value represents the equality of distribution of the binary variables (using a χ^2 test) and continuous variable (using a Kruskal-Wallis test).

Note that the subjects' choices result in only 2 categories (instead of 4): individualistic and prosocial. The proportion of egalitarians corresponds to the answers from Kahan *et al.* (2011).[12] Regarding questions from ISSP's survey (Roberts *et al.*, 2023), subjects demonstrate relative intolerance towards global inequality (Intolerance Global Inequalities) and agree that wealth differences are unfair (Fairness belief of Global Differences). Answers

[12]I assess the sum of the Likert scale score for all the questions and classify subjects as egalitarian if their sum is above the median score for questions relative to egalitarian preferences (i.e. 27.00) and hierarchical if their sum is above the median score for questions relative to hierarchical preferences (i.e. 15.00).

are more nuanced regarding the implementation of a tax in rich countries for poor countries (Global Tax Support) and the ability to work in wealthy countries for people from poor countries (Migration Support). The distributions of answers for each question are presented in Figure 2.A1 in Appendix 2.A.

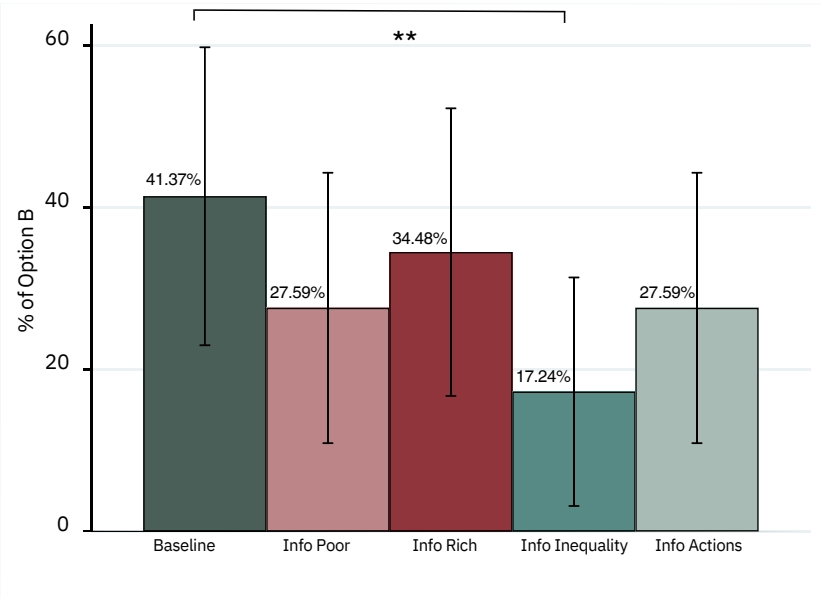
2.5.1 Willingness to move towards equality

This section investigates the influence of the treatments on the choice of Option B (equal distribution) by Player 1 in Group 1.[13] Among the whole sample, 29.65% of Player 1s opt for the equal distribution. One of the explanations for the low rate of redistribution could lie in the status quo bias (Samuelson & Zeckhauser, 1988). Since Player 1 first receives a higher allocation, it is costly to change the distribution. Figure 2.1 illustrates the proportion of Player 1s (Group 1) who choose Option B, by treatment. In Baseline, subjects seem to exhibit greater aversion to inequality. Subjects redistribute slightly but not significantly more in the baseline (41.37%) compared to the treatments (χ^2, $p = 0.12$). The richest player chooses slightly less often Option B when the poor's allocation of Group 2 is disclosed (27.59%) compared to the rich's (34.48%). However, data show no significant difference in redistribution choice between the two treatments (χ^2, $p = 0.57$) or between the two treatments and the baseline (respectively χ^2, $p = 0.27$ and $p = 0.59$).

Regarding Info Inequality (17.24%), the difference with Baseline is significant (χ^2, $p = 0.04$), meaning that dictators in Group 1 are less likely to choose an equal distribution when they observe that the distribution of Group 2 is more unequal than theirs. Overall, this first outcome supports the first hypothesis. Moreover, subjects tend to redistribute slightly less in Info Actions (27.59%) than in the baseline but the difference is not significant (χ^2, $p = 0.27$). Player 1s in Group 1 do not choose significantly more to redistribute equally when observing that Group 2 does not redistribute. Since I do not observe any significant effect of the Info Actions treatment, I reject my second hypothesis.

[13]Only one Group 2 was present in each condition and the dictator always chooses Option A.

Figure 2.1: Percentage of Option B, by treatment



Notes: Confidence intervals at 95%.

Result 1: *Hypothesis 1 is supported. Participants are significantly less willing to move towards equality when being aware of the other group's higher inequality (Info Inequality), compared to the baseline.*

Result 2: *Overall, Hypothesis 2 is not supported: the decision not to redistribute in Group 2 does not significantly influence Group 1 (Info Actions).*

Regarding Info Poor and Info Rich, participants mainly succeed in guessing the status of the player to whom the disclosed allocation belongs (Status Guess, respectively 90.00% and 80.00%).[14] Also, correctly guessing the status of Group 2's player does not significantly affect the choice to redistribute (χ^2, $p > 0.10$). In Info Actions, among all the players from Group 1, 79.31% believe the other will not redistribute equally the resources (Choice Guess).[15] Believing that Group 2 will redistribute significantly increases the likelihood that Player 1 in Group 1 chooses Option B (χ^2, $p = 0.07$). This supports the idea of the *moral wiggle room*

[14]Player 1s and 2s do not significantly differ in their beliefs regarding the status (Status Guess) of Group 2's player (χ^2, $p > 0.10$).

[15]I do not observe a significant difference between expectations of different players within the group: 82.76% of Player 1s and 75.86% of Player 2s believe that the other group will choose Option A (χ^2, $p = 0.52$).

in the decision to redistribute equally, when predicting the action of another group.

In Table 2.3, I conduct two specifications using a Logit model. The dependent variable is Choice of Option B in Group 1, which takes the value 1 if the subject chooses to redistribute equally and 0 otherwise. Specification 1 focuses on the effect of treatments and specification 2 includes control variables: Female (takes the value 1 if the subject is female), Age, Prosocial (takes the value 1 if the subject is classified as prosocial in the SVO task), and Egalitarian (takes the value 1 if the subject is classified as egalitarian using the Kahan *et al.* (2011) survey). I also control for the field of education. In all specifications, I include the value of the marginal effect of the treatment Info Inequality.

The coefficient and the marginal effect of Info Inequality in both specifications confirm previous observations. Observing greater inequality in another group reduces the probability of choosing Option B by 20% according to the marginal effect. This effect is robust to the addition of control variables. As shown by the positive and significant coefficient of Prosocial, redistribution choices are mainly driven by prosocial motivations such as the reduction of payoff difference between the two players.[16] In almost all treatments, individualists always choose Option A (unequal distribution).[17] In Baseline prosocial subjects select Option B more often compared to Option A (χ^2, $p = 0.00$). In Info Inequality, it is the opposite since they opt more for the unequal distribution than the equal one (χ^2, $p = 0.04$). This pattern is illustrated by Figure 2.A2 (Appendix 2.A). To further understand the determinants of choosing Option B in each treatment, I conduct separate logit regressions for each treatment. The results are presented in Table 2.B1 (Appendix 2.B.1).

[16] According to Murphy *et al.* (2011): "*a [decision-maker] may endeavor to maximize her own payoff (individualistic), maximize (competitive) or minimize (inequality averse) the difference between her own and the other person's payoff, or maximize joint payoffs (prosocial)*" (p.771). Further in the paper, prosocial motivations refer both to the willingness to maximize joint payoffs and to inequality aversion. Since, in this experiment, the choice of redistribution does not affect the joint payoffs, I only consider prosociality as the willingness to reduce the difference in payoff between the two players.

[17] The only exception is for Info Actions in which only one individualist chooses Option B.

Table 2.3: The impact of the treatments on the choice of allocation

	Choice of Option B	
	(1)	(2)
Info Poor	-0.62	-1.01
	(-1.10)	(-1.54)
Info Rich	-0.29	-0.66
	(-0.54)	(-1.00)
Info Inequality	-1.22**	-1.39**
	(-1.96)	(-2.04)
Info Actions	-0.62	-0.30
	(-1.10)	(-0.43)
Female		0.47
		(0.89)
Age		0.02
		(0.92)
Education		-0.06**
		(-2.19)
Prosocial		3.89***
		(3.72)
Egalitarian		0.32
		(0.71)
Constant	-0.35	-3.16***
	(-0.92)	(-2.31)
Observations	145	143
Pseudo R^2	0.03	0.27
Marginal effect of Info Inequality	-0.20**	-0.21**

Notes: Logit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-2: Choice of Option B. Independent variables described in text. *, **, *** indicate significance at the 10%, 5%, 1% level, respectively. We do not include in specification 2, two observations from participants not declaring their gender.

2.5.2 Fairness perceptions

In addition to the dictator's decision analysis, I assess the perception of fairness regarding the allocation received.[18] For the reminder, after receiving information about Group 2 and the allocation corresponding to the option chosen by the dictator, Player 2 is asked if they think the allocation received is fair or not. As shown in Table 2.2, on average 47.51% of Player 2s in Group 1 perceive the allocation received as fair. Perception of fairness is strongly correlated to the amount received (Pearson's r = 0.62, $p = 0.00$) and appears to be influenced by the type of information disclosed. Figure 2.2 illustrates the proportion of Player 2s receiving Option A who perceive the allocation as fair. In Baseline, only 5.88% of Player 2s perceive that the allocation received is fair. In Info Poor, their perception of fairness is slightly higher but not significant compared to Baseline (χ^2, $p = 0.68$). In Info Rich, the proportion of subjects perceiving the allocation as fair goes up to 57.89%, which is significantly higher than in Baseline and Info Poor (χ^2, $p = 0.00$). In this treatment, Player 2 may anticipate that Player 1 will not redistribute to preserve their wealthiest status (Ferrer-i-Carbonell, 2005; Clark & d'Ambrosio, 2015).[19] Since Player 2 is informed that both rich's player are equivalent (6 €), they may believe that they are the only one to perceive less. Player 2 anticipates that the dictator might not give them an additional payoff to remain as rich as the richest player in Group 2. Therefore, the alignment of negative beliefs with negative feedback does not generate as much disappointment as among the other Player 2s in the other conditions.[20]

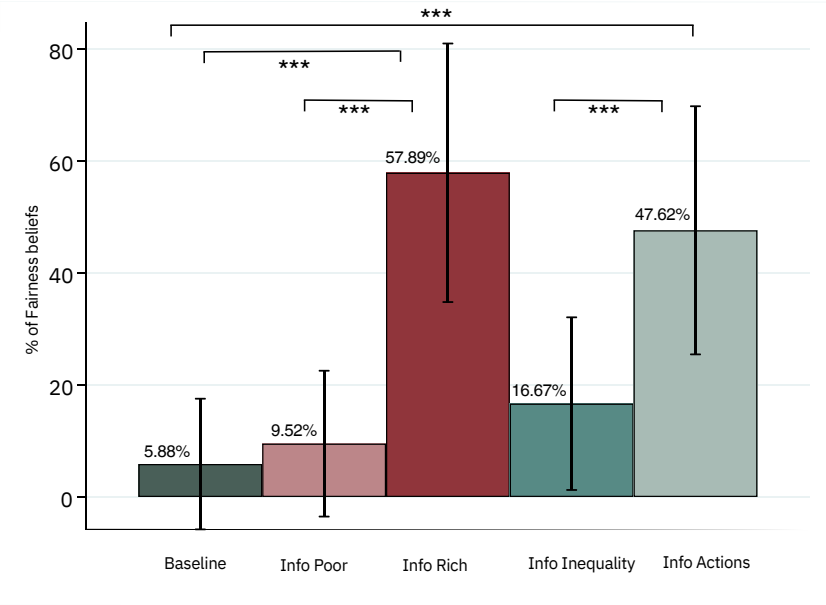
Regarding Info Inequality, fairness perception is slightly higher compared to Baseline but the difference is not significant (χ^2, $p = 0.30$). Player 2's fairness perception increases in

[18] Note that in this part I still exclude Group 2's Player 2 to avoid confusion. It is noteworthy that Player 2 in Group 2 consistently perceives the received amount as unfair, as they never experience an equitable distribution.

[19] Note that the fairness perceptions do not significantly differ when the subject correctly guesses the status of whose allocation is disclosed (Status Guess, χ^2, $p = 0.30$).

[20] To delve further into this notion, an analysis of the textual responses of participants to the question "why you believe the allocation is (un)fair?" is provided in Appendix 2.C. Overall, a significant difference in the emotions expressed is not observed based on the treatment.

Figure 2.2: Percentage of Player 2s who perceive the allocation received (Option A) as *fair*, by treatment



Notes: Confidence intervals at 95%.

Info Actions, both compared to Baseline and Info Inequality (χ^2, $p = 0.30$ and $p = 0.02$ respectively). The higher tolerance in Info Actions aligns with findings from Sanfey (2009). Player 2 in Group 1 previously witnessed Group 2's decision to not redistribute equally. This may lead them to form anticipations of inequality and since the feedback confirms their prediction, they do not feel disappointed (Gill & Prowse, 2012). However, data cannot confirm this pattern, as Player 2s do not express significantly higher perceptions of fairness when they expect Group 2's dictator not to redistribute, compared to when they expect Group 2's richest redistributes (χ^2, $p = 0.43$). The lack of significance in Info Inequality and the significant impact of Info Actions on Player 2's fairness perception partially supported the third hypothesis. To further understand the fairness perception among those who receive an unequal allocation, I conduct logit regressions with Fairness Perception (takes the value 1 if the subject perceives the allocation as fair and 0 otherwise) as the dependent variable. The results are presented in Table 2.B2 (Appendix 2.B.2).

Result 3: *Hypothesis 3 is partially supported. Player 2's perception of fairness increases*

when they know the wealthiest's payoff in the other group (Info Rich) or if the other group does not redistribute (Info Actions). However, information about other group's level of inequality (Info Inequality) does not influence the recipient's perception of fairness.

2.6 Discussion

The primary finding reported in the chapter is that the wealthiest individuals are less inclined to achieve a more equal situation in the presence of higher inequality comparison. This result raises questions about the pursuit of equality and this section explores the plausible mechanisms driving this behavior.

First, this study does not claim to alter inequality aversion *per se* and it explains the use of the expression "willingness to move towards equality/equitable situation" instead. However, this is precisely the definition of inequality aversion in Fehr & Schmidt (1999). According to Fehr & Schmidt (1999), inequality aversion is defined as the willingness to "*give up some material payoff to move in the direction of more equitable outcomes*" (p.819). Individuals who choose not to redistribute might still be averse to inequality but the observation of *worse* situation in another group may lead them to minimize the need for redistribution in their own group. This is particularly concerning for the pursuit of equality or the acceptability of redistributive policies for instance. It echoes research demonstrating that citizens may prioritize certain groups or issues as more deserving than others (Laenen & Roosma, 2022).

Second, the main explanation for the observed behavior could lie in the status quo bias (Samuelson & Zeckhauser, 1988) which is intensified by the presence of a comparison element. Wealthiest individuals may compare themselves with *worse* situations to justify maintaining the status quo that already gives them the illusion of superiority. Based on the Fehr & Schmidt (1999)'s inequality aversion model, in this experiment, Option A provides a lower disutility derived from advantageous inequality to Player 1 in Group 1 compared to Group 2 since the difference in the initial allocations of Player 1 and 2 is smaller in Group 1. If the

individual is averse to inequality, the comparison with Group 2 should satisfy them because they perceive their group as less unequal. What remains concerning is the satisfaction of being slightly better rather than striving for complete equality. A person who is genuinely averse to inequality, in the sense of Fehr & Schmidt (1999), should seek to reduce inequality regardless of the context. Moreover, the costs of achieving equality are lower for the decision-maker in the less unequal group, compared to the most unequal one. An interesting extension to conduct would be to examine dictator's decision of redistribution when observing a more equal group.

Third, contrary to the second hypothesis, I do not observe a significant effect from disclosing information about another group's decision not to redistribute. This result should be interpreted with caution, as only one Group 2 was present in each condition. Since this group chose not to redistribute, the analysis of a potential mimicry effect is limited (Cason & Mui, 1998; Goeschl *et al.*, 2018). Additional observations are required to understand the mechanisms behind the expectations of actions from the less unequal group and more precisely, when the group anticipates that the most unequal (would not) redistribute and in fact observes the opposite.

2.7 Conclusions

This chapter investigates whether rich members of a group of two players seek equality within their group when 1- observing a more unequal group than theirs, and 2- being informed that a more unequal group that theirs decides not to redistribute equally. Subjects primarily experience inequality and the richest player has the possibility to suppress inequality within their group. One notable finding is that only a few dictators opt for equitable redistribution and the rest favors maintaining the status quo instead. Wealthiest subjects appear to act more indifferent when aware of more pronounced inequality in another group. The first hypothesis is supported and this outcome provides evidence that people tend to minimize inequality

when comparing with *worse*. This result is particularly concerning because it reveals that subjects perceive their group's inequality as less severe when compared to another group's situation. Since the perception influences Player 1's actions to restore equality within their group, it demonstrates that social comparison compromises the process of achieving equality.

It is worth noting that subjects significantly choose to redistribute less only in Info Inequality. The lack of significance in Info Rich and Info Poor suggests that Player 1 does not opt for Option A (unequal distribution) more frequently to preserve their wealthiest rank. Indeed, by choosing to redistribute less in Info Rich, this result would have provided evidence that wealthy individuals are unwilling to act for equality if it possibly compromises their higher status.[21] Additionally, the non-significant impact of Info Actions indicates that the wealthiest does not significantly adjust their redistribution choices in response to another group's decision not to redistribute.

Another notable outcome is the reversal of behavior among pro-social individuals. Since the order of the tasks was not randomized, I cannot rule out the possibility that subjects choose more prosocial outcomes in the SVO task to compensate for their selfish choice in the DG task. This effect is likely to occur in all treatments and the proportion of prosocial individuals between the baseline and Info Inequality is similar (χ^2, $p = 0.12$). Nevertheless, due to the significant impact of Info Inequality on redistributive choice, subjects might be more inclined to opt for altruistic choices in the non-incentivized SVO task, possibly as a form of compensation.

Finally, this work paves the way for other possible projects, including repeated settings where inequality fluctuates over time. Future experiments could involve varying the distance between the decision-maker and the receiver, perhaps by allowing the former to justify their decisions. The notion of distance is an important issue to address because the effects of comparison and the perception of inequality depend on the proximity to the reference group (Hvidberg *et al.*, 2020; Xu *et al.*, 2023). On the other hand, the tolerance of inequality

[21] Note that their higher status is not compromised in Info Poor since the poorest in Group 2 is still poorer than Player 1 in Group 1, even when choosing Option B.

by less advantaged individuals appears to be shaped by awareness of others' actions and the wealth of the richest in another group. In this chapter, the tolerance of the poorest is assessed through the declarative way. Further investigations might allow the disadvantaged to actively restore equality, such as by burning money from the richest player for instance.

2.A Descriptive statistics

Figure 2.A1: Distribution of answers to ISSP's survey

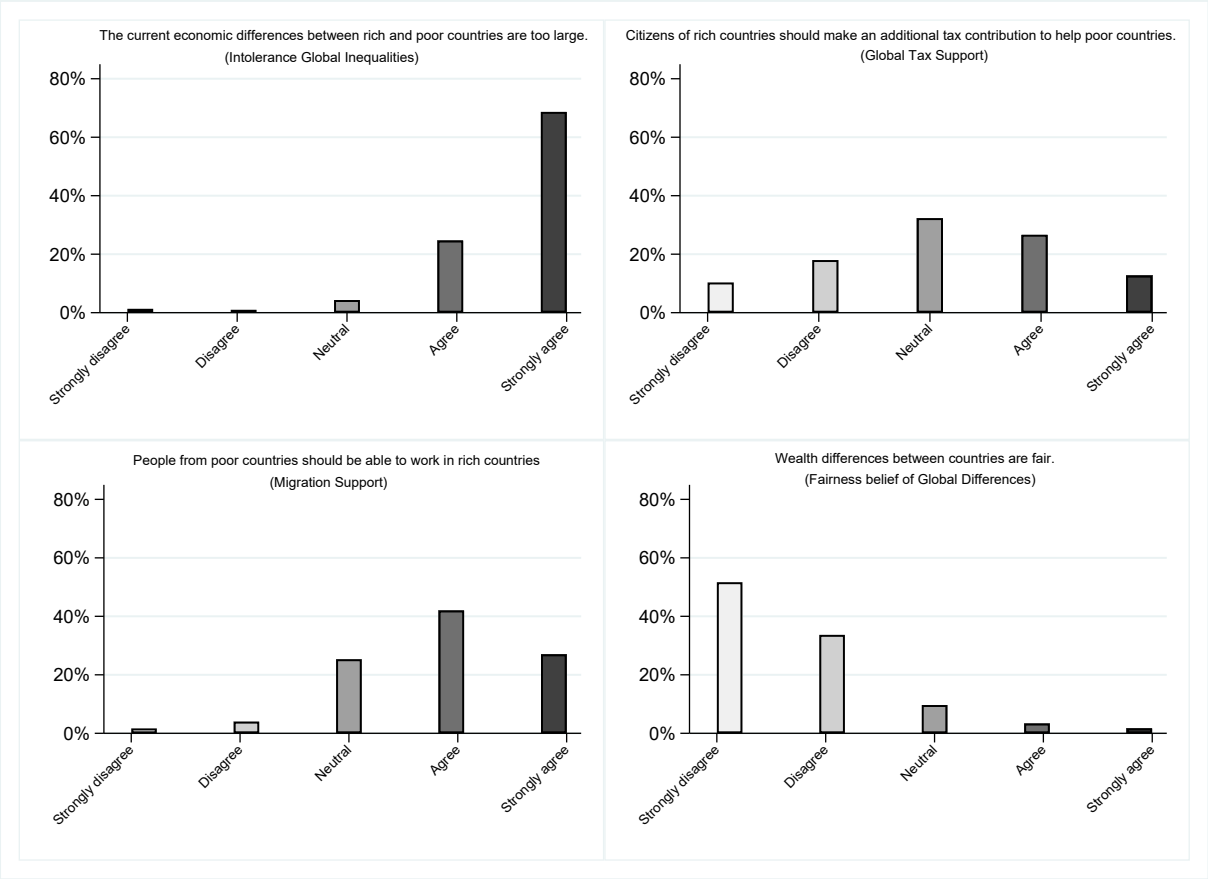
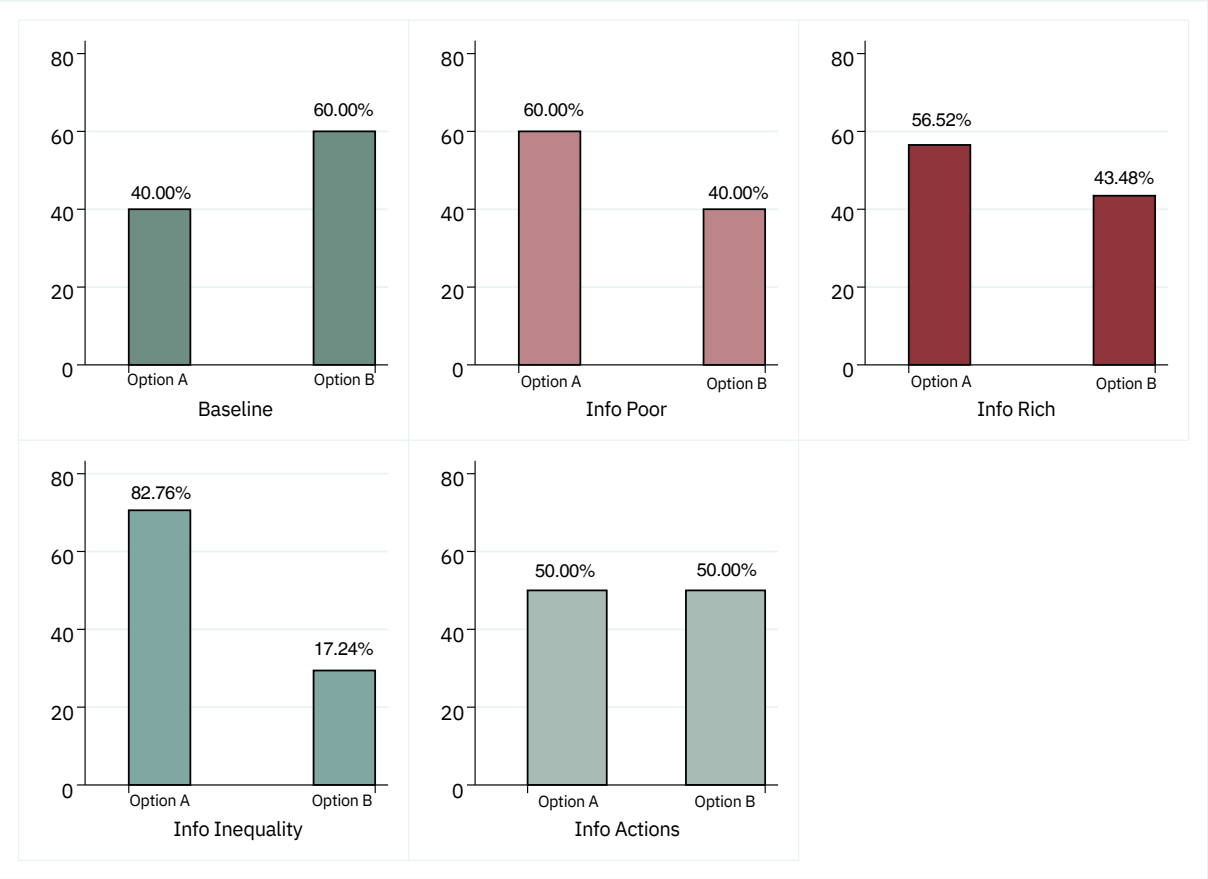


Figure 2.A2: Distribution of choice among prosocial subjects, by treatment



Notes: Prosociality is captured by the SVO task (Murphy *et al.*, 2011).

2.B Additionnal tables

2.B.1 Willingness to move towards equality

Table 2.B1 presents five specifications with Choice of Option B (in Group 1) as the dependent variable and the variables of interest include sociodemographics (Female, Age, and Education) as well as control variables: Egalitarian and the Likert scale scores of the 4 questions from ISSP's survey. For Info Poor and Info Rich (specifications 2-3), I incorporate Status Guess: takes the value 1 if the subject correctly guesses that the disclosed allocation belongs to the poorest (richest) player, and 0 otherwise. Specification 5 contains Choice Guess, which takes the value 1 if Group 1 believes that Group 2 would redistribute and 0 otherwise. Note that Prosocial is not included in the specifications since every individualist behaves in the same way in almost every treatment (except in Info Actions but it represents only one observation). Moreover, in specification 4, Female and Intolerance Global Inequalities are omitted since no men choose the equal distribution and all subjects who choose the equal distribution strongly agree regarding global differences between poor and rich countries. The same lack of variability occurs for Education, for which arts and economics students are the only ones to choose to redistribute.

As shown by the non-significance of the coefficients in almost all specifications, the measured variables are not capable of explaining the decision to redistribute within the group. Based on the chi-square value, all the models are non-significant ($p > 0.10$), except for specifications 4 and 5 ($p = 0.01$ and $p = 0.02$ respectively). Given the positive and significant coefficient of Egalitarian in specification 4, egalitarianism significantly and positively influences the willingness to restore equality. Subjects that are classified as egalitarian following Kahan *et al.* (2011)'s survey are 55.25% more prone to redistribute equally the resources ($p = 0.00$).

Table 2.B1: Determinants of choice of Option B in each treatment

	Choice of Option B				
	Baseline	Info Poor	Info Rich	Info Inequality	Info Actions
	(1)	(2)	(3)	(4)	(5)
Female	2.97	-2.40	1.49		0.47
	(1.62)	(-1.27)	(0.73)		(0.29)
Age	-0.25	-0.00	0.06	-0.61	0.10*
	(-1.09)	(-0.07)	(0.91)	(-1.01)	(2.01)
Education	0.18	0.51	-0.59		0.01
	(1.07)	(1.54)	(-1.61)		(0.08)
Egalitarian	0.63	-0.49	2.66	8.91***	1.34
	(0.52)	(-0.34)	(0.99)	(3.36)	(0.70)
Intolerance Global Inequalities	-1.60	0.23	1.22		-0.98
	(-1.30)	(0.15)	(0.81)		(-1.21)
Global Tax Support	-0.22	1.31	-0.29	1.74	-0.85
	(-0.30)	(1.31)	(-0.56)	(1.43)	(-1.58)
Migration Support	0.91	3.07*	-2.46	-7.48***	-0.67
	(1.45)	(1.86)	(-1.09)	(-3.88)	(-1.19)
Fairness belief of Global Differences	1.79	0.24	-2.54	-1.26	-1.12
	(1.56)	(0.15)	(-1.29)	(-1.10)	(-1.48)
Status Guess		0.45	-3.87		
		(0.34)	(-1.16)		
Choice Guess					1.61
					(1.43)
Constant	3.35	-21.00	11.67	29.49*	6.51
	(0.42)	(-1.62)	(1.33)	(1.95)	(1.24)
Observations	29	29	29	29	29
Pseudo R^2	0.30	0.34	0.47	0.57	0.29
Prob $> \chi^2$	0.44	0.61	0.83	0.01	0.02

Notes: Logit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-5: Choice of Option B. Independent variables described in text. *, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

2.B.2 Fairness perceptions

Table 2.B2 presents two specifications using a logit model, with Fairness Perception (takes the value 1 if the subject perceives the allocation as fair and 0 otherwise) as the dependent variable. Column 1 assesses the effect of the treatments and column 2 incorporates sociodemographics (Female, Age, and Education) and control variables (Prosocial and Egalitarian). As shown by the significant and positive coefficient in both specifications, both Info Rich and Info Actions increase Player 2's perception of fairness. This result holds with the addition of control variables. The marginal effect of Info Rich is stronger compared to Info Actions. The disclosure of the wealthiest player enhances the perception of fairness by more than 50% against approximately 40% for Info Actions.

Table 2.B2: The impact of the treatments on the fairness perceptions

	Fairness Perception	
	(1)	(2)
Info Poor	0.52	0.54
	(0.41)	(0.42)
Info Rich	3.09***	3.48***
	(2.72)	(3.10)
Info Inequality	1.16	1.40
	(0.99)	(1.14)
Info Actions	2.68**	2.83**
	(2.38)	(2.36)
Female		0.16
		(0.34)
Age		-0.03
		(-0.83)
Prosocial		-0.79
		(-1.47)
Egalitarian		-1.02
		(-1.64)
Education		-0.03
		(-0.39)
Constant	-2.77***	-1.36
	(-2.68)	(-1.00)
Observations	102	102
Pseudo R^2	0.19	0.26
Marginal effect of Info Rich	0.52***	0.54***
Marginal effect of Info Actions	0.42***	0.40***

Notes: Logit model regressions. t statistics in parentheses and standard errors are robust. Dependent variable in 1-2: Fairness Perception. Independent variables described in text. *, **, *** indicate significance at the 10%, 5%, 1% level, respectively. Note that 1 subject that did not want to declare their gender was assigned to Player 2. In the table, I only consider those who receive an unequal allocation.

2.C Additional analysis: Participants' answers

The post-experimental content analysis of Players 2's responses was conducted by two independent research assistants who were blind to the hypotheses. They were asked to classify responses into six categories: neutral, description of the other's moral values (i.e. Player 1), spitefulness, thankfulness, speculation/expression of own moral position and pragmatic. The frequency of the responses for each category is summarized in Table 2.C1 and I distinguish between *Panel A* (those who receive an unequal distribution) and *Panel B* (those who receive an equal distribution). Note that for some answers the research assistants attribute several categories. The answers could thus be at the same time in different categories. I only exclude some answers of "neutral" when they are also matched with another category. Figure 2.C1 illustrates the distribution of the categories, based on the treatment, and distinguishes by the panel.

Neutral: The two research assistants significantly differ in their coding of the "neutral" answers (χ^2, $p = 0.00$). They however find a similar pattern regarding the proportion of neutral answers based on the allocation received. Players experiencing inequality (*Panel A*) are significantly more likely to express emotion in their answers, compared to *Panel B* (χ^2, $p = 0.00$). An example of an answer that both research assistants agree on: "*It is an equitable decision*".

Description of the other's moral values: For this category, the coding of the two research assistants does not significantly differ (χ^2, $p > 0.10$). *Panel A*'s responses contain significantly more a vocabulary referring to the moral values of their partner, compared to those who have received an equal offer (χ^2, $p = 0.06$ for researcher 1 and $p = 0.02$ for researcher 2). This category refers to participants who mentioned either negative (selfishness) or positive (generosity) moral values exhibited by their partner. For instance, one participant wrote: "*(...) so if she/he is a good person, she/he should have distributed it equally.*"

Spitefulness: The category referring to spitefulness appears very similar to the former

one, which may explain why the responses of researchers significantly differ (χ^2, $p = 0.00$). Moreover, since the two researchers did not have access to the topic or the result of the study, some answers may not be interpreted correctly.[22] *Panel A* is again more likely to express spitefulness than *Panel B*. In this experiment, Player 2 cannot take their revenge with actions but their spiteful feeling could be illustrated in their answers with the use of insult for example. For instance: "*(...) it shows the lack of empathy and greed for money*".

Thankfulness: As opposed to resentment, subjects may express gratitude. The research assistants do not agree on the classification (χ^2, $p = 0.00$) but they both find that *Panel B* express more gratitude than *Panel A*. The gratitude is expressed by positive words such as "generosity" or by the use of "thanks". For instance, both researchers agree to classify the following sentence as thankfulness: "*[she/he] has decided to be fair, thank you*".

Speculation, expression of own moral position: The classification between the two researchers aligns. Player 2s in *Panel A* are significantly more prone to express their own moral position compared to *Panel B* (χ^2, $p = 0.05$ for researcher 1 and $p = 0.03$ for researcher 2). Note that I voluntarily use the word *speculation*. Given the low proportion of subjects who choose the equal distribution it is very unlikely that all of those who declare they would choose the equal distribution would actually have done so. For instance, one participant who receives an unequal distribution, declared: "*if I were player 1, I would think of the welfare of both of us*".

Pragmatic: Finally, some subjects express a sense of pragmatism in their answer. The two research assistants do not agree on the frequency (χ^2, $p = 0.00$) but their coding reflects the same trend: the frequency of pragmatic answers does not significantly differ based on the allocation received (χ^2, $p > 0.10$). Pragmatism corresponds to answers that "excuse" the selfishness of the dictator or that may explain the decision by the nature of the game. For instance, one participant (from *Panel B*) answered: "*the decision seems fair to me as we both had the possibility to be player 1 or 2*". Another participant (from *Panel A*) expressed: "*It's*

[22]For instance, "*partially equitable*" was considered spiteful while the player received an equal offer.

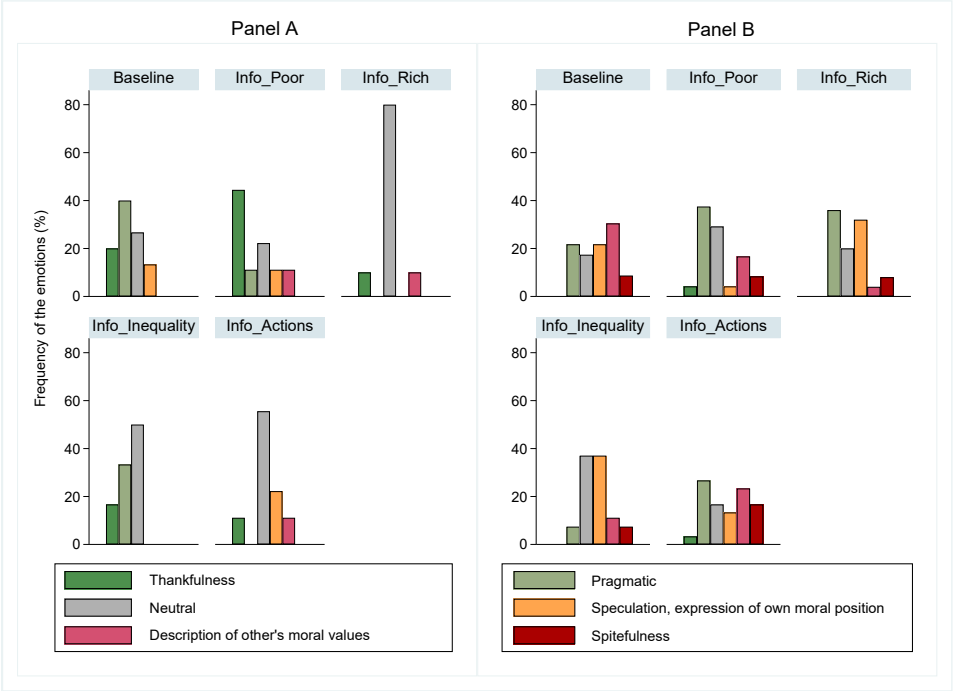
not fair because fairness implies equality, but it's well done, I would have done the same".

Table 2.C1: Answers to the question "Why do you think the decision is (un)fair?"

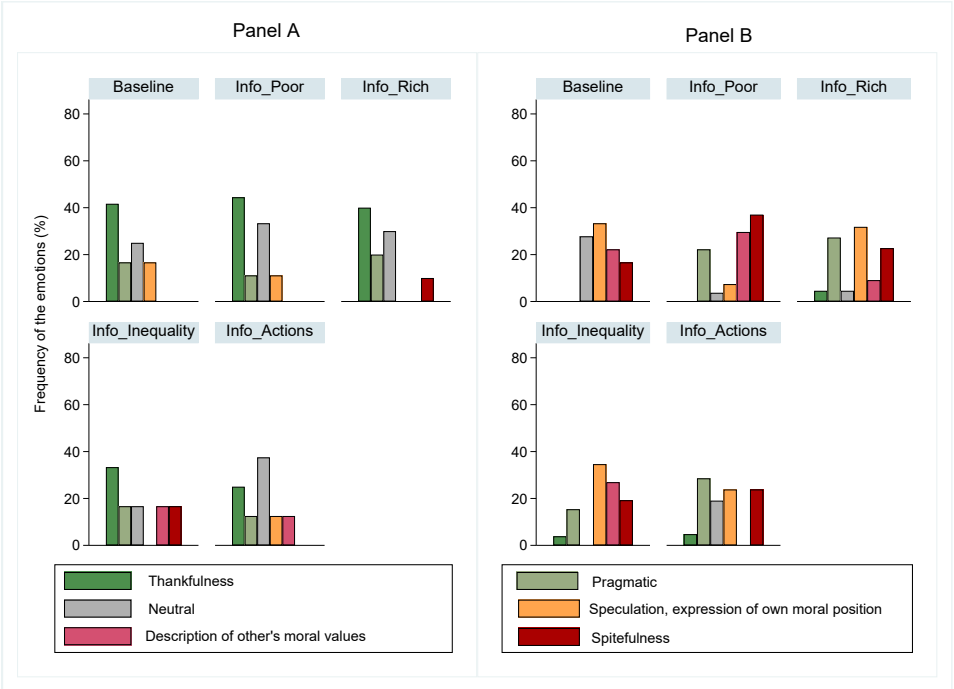
	Frequency (%)	
	Research Assistant 1	Research Assistant 2
Neutral	35.33	16.00
Description of other's moral values	16.67	15.33
Spitefulness	8.67	20.00
Thankfulness	8.00	13.33
Speculation, expression of own moral position	22.00	22.00
Pragmatic	28.00	19.33
Panel A: Those who receive an unequal distribution		
Neutral	28.97	10.28
Description of other's moral values	20.56	19.63
Spitefulness	12.15	26.17
Thankfulness	1.87	2.80
Speculation, expression of own moral position	26.17	27.10
Pragmatic	30.84	16.28
Panel B: Those who receive an equal distribution		
Neutral	51.16	30.23
Description of other's moral values	6.98	4.65
Spitefulness	0.00	4.65
Thankfulness	23.26	39.53
Speculation, expression of own moral position	11.63	9.30
Pragmatic	20.93	16.28

Figure 2.C1: Distribution of the categories, by treatment, according to the research assistant

(a) Research assistant 1



(b) Research assistant 2



2.D Instructions (translated from Spanish)

Welcome to this experiment! This experiment consists of three parts. In the first part (**Part I**), you will be matched with another player in the room, and each of you will receive an initial endowment. Depending on your role in the group, you will have the option to redistribute the endowments of both players:

- If you are randomly chosen as **Player 1**, you will have the option to redistribute the endowments of both players.
- If you are randomly chosen as **Player 2**, you will have a passive role and observe the decisions of **Player 1**. Therefore, you cannot determine the payments for either yourself or the other player.

[$Info_Poor$ & $Info_Rich$:
Before you or the other player in your group make the decision, you will observe the initial endowment of another player from another group in the session.]

[$Info_Inequality$:
Before you or the other player in your group make the decision, you will observe the initial endowments of both Player 1 and Player 2 from another group in the session.]

[$Info_Actions$:
Before you or the other player in your group make the decision, you will observe the decision made by Player 1 from another group in the session.]

In the second part (**Part II**), you will play a short game, independent of the first part. In the third part (**Part III**), you will be asked to answer a questionnaire. Your final payoff corresponds to the payoff receive in **Part I** plus the participation fee of 3 €.

..... *Next Page* .....

[Group 1 :

Player 1 : You have been randomly chosen as **Player 1**. You receive an endowment of 6 €, while the other player receives an endowment of 3 €. Next, you will have the option to change your earnings and those of Player 2.

Player 2 : You have been randomly chosen as **Player 2**. You receive an endowment of 3 €, while the other player receives an endowment of 6 €. Next, the other player will have the option to change your earnings and theirs.]

[Group 2 :

Player 1 : You have been randomly chosen as **Player 1**. You receive an endowment of 6 €, while the other player receives an endowment of 1 €. Next, you will have the option to change your earnings and those of Player 2.

Player 2 : You have been randomly chosen as **Player 2**. You receive an endowment of 1 €, while the other player receives an endowment of 6 €. Next, the other player will have the option to change your earnings and theirs.]

..... *Next Page* .....

[$Info_Poor$: For your information, in another group in the session, one of the players received an initial endowment of **1 €** (3 €). According to you, this endowment corresponds to the one of :

- The poorest
- The richest]

[$Info_Rich$: For your information, in another group in the session, one of the players

received an initial endowment of **6 €**. According to you, this endowment corresponds to the one of :

- The poorest
- The richest]

[$Info_Inequality$: For your information, in another group in the session, Player 1 received an initial endowment of **6 €** and Player 2 received an initial endowment of **1 €** (3 €).]

[Group 1 | $Info_Actions$: In another group, Player 1 received an initial endowment of **6 €** and Player 2 received **1 €**. Player 1 can give **3.50 €** to Player 2 and keep **3.50 €** for themself or not change anything. What do you think the other group will do:

- Not change the allocations
- Give 3.50 € to Player 2 and keep 3.50 € for themself]

.....*Next Page*.....

[Group 1 | $Info_Actions$:

Player 1 decided to give **1 €** (3.50) to Player 2 and keep **6 €** (3.50) for themself, instead of giving **3.50 €** (1) to Player 2 and keeping **3.50 €** (6) for themself.]

.....*Next Page*.....

NB: The order of the sentence was randomized.

[Group 1 :

Player 1 : Choose the favorite option:

- 6 € for you | 3 € for the other player
- 4.50 for you | 4.50 for the other player]

[Group 2 :
Player 1 : Choose the favorite option:

- 6 € for you | 1 € for the other player
- 3.50 for you | 3.50 for the other player]

..... *Next Page* .....

[Player 2: Player 1 decided to give you **1/3 euro(s)** (3.50/4.50) and keep **6 €**, instead of giving you **3.50/4.50 €** (1/3) and keeping **3.50/4.50 €** (6) for themself.

Do you think the decision made by Player 1 is fair?

- Yes
- No

Why ? _____]

..... *Next Page* .....

[Player 1: You have decided to give **x euro(s)** to Player 2 and keep **x €** for yourself, instead of giving them **x euro(s)** and keeping **x €** for yourself. Your payoff for this part is x €.]

[Player 2: Your payoff for this part is x €.]

..... *Next Page* .....

[Social Value Orientation Test]

..... *Next Page* .....

[Strongly Disagree | Disagree | Neither Agree nor Disagree | Agree | Strongly Agree]

[ISSP's survey]

- The current economic differences between rich and poor countries are too large.
- Citizens of rich countries should make an additional tax contribution to help poor countries.
- People from poor countries should be able to work in rich countries.
- Wealth differences between countries are fair.

[Cultural worldviews]

- We have gone too far in promoting equal rights in this country.
- Our society would be better off if wealth distribution were more equitable.
- We need to drastically reduce inequalities between the rich and the poor, and between men and women.
- Discrimination against minorities remains a very serious problem in our society.
- It seems that women, homosexuals, and other groups do not want equal rights, they want special rights just for them.
- Society as a whole has become too soft and feminine.

. . . . . *Next Page* . . . . .

- How old are you? _____
- What is your gender?
    - o Male
    - o Female
    - o Other
- What is you field of study? _____

Expérience sur le terrain

Chapter 3

The effects of reminders on engagement and walking: Evidence from a large scale experiment [1]

Every step you take. Every connection you make. I'll be watching you.

[1] This chapter is a joint work with Beatrice Braut (Department of Economics, University of Genoa, Italy) and Sarah Zaccagni (Department of Economics and Business Economics, Aarhus University, Denmark).

3.1 Introduction

Regular exercise leads to benefits that are well known: according to World Health Organization[2], it prevents and helps manage cardiovascular diseases, diabetes, and cancer. Being insufficiently active leads to a 20% to 30% increased risk of death. Moreover, physical activity has a positive effect in reducing symptoms of stress and depression and in general enhances mental well-being and cognitive abilities (Sharon-David & Tenenbaum, 2017; Matzer *et al.*, 2018). Exercising also boosts motivation for other health behaviors like eating self-regulation (Baker & Brownell, 2000; Mata *et al.*, 2009), contributing directly and indirectly in contrasting obesity. Despite the numerous benefits of engaging in regular physical activity, people often refrain from doing it. Indeed, the benefits from exercising are postponed to the future, while the costs encumber on the present. This represents a perfect application of present bias (O'Donoghue & Rabin, 1999), where individuals prioritize immediate rewards over future payoffs. In addition, in many cases to start exercising requires to pay a subscription or invest in specific equipment, on top of the time and non-monetary effort invested. In such a context, finding ways to encourage people to engage in physical activity is crucial.

In this chapter, we test in a field experiment the effectiveness of framed messages in promoting steps tracking and walking. Walking represents one of the most popular ways to be physical active and can be done at any level by everybody, without incurring any monetary cost. Moreover, as a zero-emission mean of transportation, favoring walking as an alternative to car usage is also relevant in the mobility transition.[3] Although there is no unambiguous agreement on the right number of daily steps[4], setting a daily goal and counting the steps help to build awareness on the importance of not being sedentary (Walsh *et al.*, 2016). The possibility to track, check and compare health parameters through apps, which is especially common for steps, opens the way for behavioral interventions aiming at improving health habits. These are even more needed in a context where the number of

[2]See WHO website: https://www.who.int/news-room/fact-sheets/detail/physical-activity

[3]See European Commission website: $https://transport.ec.europa.eu/news-events/news/efficient-and-green-mobility-2021-12-14_en$

[4]The New York Times, 2021: https://www.nytimes.com/2021/07/06/well/move/10000-steps-health.html

steps is dropping (Desine *et al.*, 2023). In this study, we design and implement a behavioral intervention in collaboration with a step-tracking app, *We Ward*, on a large sample of the French population. *We Ward* is a popular application to track steps, born in France and then diffused in other eight countries. What attracts its more than 20 million users is the fact that they are rewarded for their steps. However, the monetary incentive is not the only feature the app uses to promote walking habits. In the app, subjects can find information on the many benefits of walking and related issues, incur in social comparison, receive reminders and participate in walk-based games and challenges.

Apps represent a new opportunity when traditional information campaigns fail promoting good health behaviors (Tufano & Karras, 2005). Gamification is one of the most valuable tools that apps can use to influence health behaviors (King *et al.*, 2013). Messages and support through mobile devices have been found to be effective to address health behaviors, as for example increasing the take up of medications and quitting smoking (Free *et al.*, 2009; Lester *et al.*, 2010). Another useful tool that apps can exploit is the possibility to create communities, and leverage peer effects and peer comparison to reach their goal. Indeed, social factors significantly impact many health behaviors, like eating and smoking (Christakis & Fowler, 2007; Centola, 2013). Also for physical activity, it has been found that the virtual support of other athletes promotes better performance (Franken *et al.*, 2023). This also explains why posts on the ever growing sports social communities receive higher attention than posts on other social media.[5] Participants in virtual sport communities, where details of their sports activity are shared, are motivated as a result of their engagement with the app (Russell *et al.*, 2023).

Apps have a good potential for promoting health behaviors and several of them already integrate behavioral features to push a change in habits. Messages are used widely to inform and remind about good health behaviors. Several studies report them to be effective in improving health-related behaviors. However, most studies rely on self-reported measures. This is the case of papers that analyse, for example, diet choices (Freyne *et al.*, 2017), smoking

[5]See https://business.strava.com/blog/strava-clubs-vs-traditional-social-media

cessation (Rodgers *et al.*, 2005), physical activity of women after giving birth (Fjeldsoe *et al.*, 2010), compliance with testing (Nyatsanza *et al.*, 2016), adherence to medication guidelines (Vervloet *et al.*, 2012), and to vaccination plans (Busso *et al.*, 2015; Chen *et al.*, 2023).

In our experiment, we collaborate with the app *WeWard* to design and test the effect of different types of messages and different time extensions of the intervention in a setting where we can measure walking directly and without generating experimental demand effects. While there is evidence that messages and reminders can be effective in the short run and while in place (Slaunwhite *et al.*, 2009), to transform the short-term effect into a long-term behavioral change, it is necessary to lead to the formation of new habits and this process might require a more sophisticated approach. In the context of exercising combining different tools has been found to be an effective strategy (Royer *et al.*, 2015; Patel *et al.*, 2016). We introduce messages leveraging different behavioral features to amplify the positive effect of a reminder. The messages leverage respectively: sunk costs fallacy, peer comparison, and self comparison. Moreover, we study the effect of our messages in two different time frames: a shorter one of one week, and a more extended one of three weeks. We compare our messages to a basic reminder to test if it is possible to improve the effectiveness of messages, looking for an easy scalable tool to improve exercising and awareness. Our experimental design and its peculiar setting offer several advantages: the possibility to evaluate the impact on actual behavior and to track a large sample of individuals, who did not self-select into the intervention group. Our evaluation also benefits from observations recorded in the two weeks prior to the experiment and gives us the opportunity to follow-up with the subjects afterward. Using the observations collected *ex-ante*, we define different categories of walkers and determine the effect of the treatment on different subgroups of the population, following the approach of Charness & Gneezy (2009). We also have post-intervention observations to test whether the effects, if present, persist also once the message is removed.

Our findings reveal that the treatment leveraging peer comparison increases the probability of subjects opening the app compared to the standard message while the treatment leveraging self comparison is the most successful in increasing the level of activity on the app.

However, none of these effects translates into in an increase in the number of steps subjects walk. Steps follow their trend, that varies by subject and is strongly influenced by external factors as the day of the week and the holidays. Customized messages do not influence steps, no matter the duration of the treatment. When we look at heterogeneity by the initial level of walking, we find that inactive users are the only ones positively affected by the treatment, and this effect persists over time. However, the type of customized messages does not make a difference.

The remaining of the chapter is structured as follows. Section 3.2 presents our design and Section 3.3 describes our hypotheses. In Section 3.4, we outline our estimation strategy. The results are presented in Section 3.5, and we summarize our conclusions in Section 3.6.

3.2 Experimental Design

Our experiment is a between-subject design that compares the effect of different types of messages and different treatment duration on subgroups of a representative sample of *WeWard* users in France. In the following, first we lay out the design of the mobile app. Then, we describe the procedure and the sample of the experiment, and finally the treatments.

3.2.1 The step-tracking app

Launched in 2019, *WeWard* is an app whose aim is to motivate people to walk.[6] *WeWard*, diffently to other step-tracking apps, makes use of financial incentives: users earn points (*Wards*) for reaching daily step goals and these points can be converted, after reaching a certain amount, into money or used to obtain discounts on online purchases. Payment per 10,000 steps, which represents a commonly suggested goal, is equal to 0.05 €. The app gives the opportunity to earn up to 3.75 € per month. The payment is delayed until the moment

[6]For a screenshot of the homepage of the app refer to Appendix 3.A, Figure 3.A1.

the user archives a certain number of *Wards*, corresponding to 15 €[7]. More than 10,000 million users downloaded the app in France and Italy and it has recently expanded to other countries such as Spain, Germany, the United States, Belgium, the United Kingdom, Japan[8], and the Netherlands. According to the company, the app works well as it has been found to increase the number of steps by 24%.

WeWard has many features that contribute to explaining its success, beyond monetary incentives. The first is gamification: the app features and design allow to transform a physical activity into a game. For example, it offers challenges where users can earn a badge by completing a walk in a certain amount of time or at a certain time or for walking a certain amount of steps for N consecutive days. The mechanism behind *WeWard* implies the generation of sunk costs as the steps done are not immediately convertible, thus the users can perceive their initial steps as an investment for future returns. These sunk costs can act as a self-management device (Hong *et al.*, 2019): subjects are induced not to give up because they have already invested in their walking with the scope of obtaining a reward. The app also keeps track of individual progress, allowing subjects to compare one's own performance with previous data and checking their progresses. The self comparison can boost individuals' motivation to improve their performance over time, such to create new walking habits. Keeping track of own parameters can in some cases act as motivation to exercise (De La Torre *et al.*, 2021). The app gives also the opportunity to connect to other subjects and create a sense of community between the users. It also allows to compare subjects' performance with others when added in their network. Comparison with peers can influence the subject's behavior and push them to accumulate more steps and use the app more, also according to the relative position between own and peer's performance (Franken *et al.*, 2023). The incentives that the app provides consists in points, called *Wards*, that can be collected when the subject reaches one of the 6 thresholds corresponding to 1,000, 3,000, 6,500, 10,000, 15,000, and 20,000 steps a day. The number of *Wards* earned increases not proportionally with the number of steps

[7]Payment details refer to the time of the experiment and were constant throughout the period of experimentation.

[8]Last update: March 2024.

and steps between two thresholds do not give any remuneration. Importantly, subjects do not only need to walk these amount of steps but also to validate them by opening the app and clicking on the number of steps that acts as validation button. This technical feature allows to keep high the attention of subjects on their daily steps: it is not sufficient for a subject to be a good walker but has also to be conscious that they are walking enough or not.

Overall, the *WeWard* app can effectively contribute to the formation of walking habits and to change them. Moreover, it can also change the attention and importance given to this very common but maybe underestimated physical activity.

3.2.2 Procedure and Sample

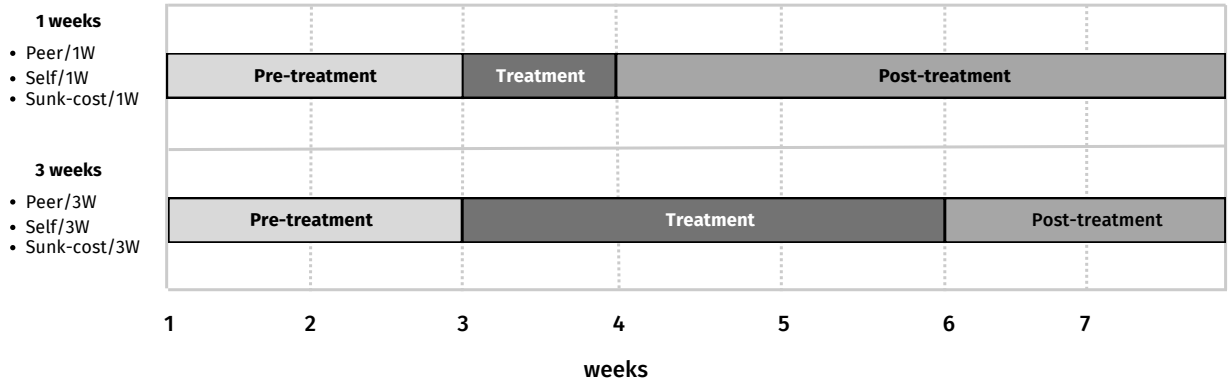
To implement our field experiment, we collaborated with the start-up *WeWard*. This collaboration had two main advantages. First, it allowed us to run a large-scale experiment with a sample of more than 20,000 subjects. Second, since *WeWard* directly tracks the walking data of its users, implementing our intervention on this app gave us the opportunity to measure actual behavior in the everyday life of unaware subjects, thus allowing to get rid of self reporting biases and demand effects.

The study runs between December 2022 and January 2023, for a total of seven weeks. We started collecting two weeks of observations as the baseline. In these two weeks, subjects were not exposed to the treatment and only received the standard reminder that the app has always been sending. Starting from the third week we introduced the treatments: one of our three messages replaced the basic reminder in the treated sample while the message was not changed for the control group.[9] The treatments were in place for one week, or for three weeks. After the treatment period, we collected post-treatment observations for four weeks for the ones assigned to the shorter duration of the treatment, or for two weeks for the ones assigned to the longer one. The total duration of the study was of seven weeks for all

[9]For a description of the treatments, see Subsection 3.2.3.

subjects. Figure 3.1 shows the timeline of the experiment. The experiment was pre-registered at the AEA RCT Registry.[10]

Figure 3.1: Timeline of the project



A total of 20,187 individuals were included in the sample, which corresponds to 422,076 observations. Our sample is representative of *WeWard* users and restricted to subjects who, when signing up, expressed consent to receive notifications from the app. This was a necessary condition for the implementation of the intervention through users' mobile phones. Moreover, since the app allows to change the time when the reminder is received, the selection of the sample was made between the ones for which the reminder was set at the default time (9pm), which represents the most numerous group of users. These users therefore are the ones who did not opt-out from reminders and also did not took the time to personalize its timing, therefore we can think they do not have a specific interest in the reminder. In our data, 52.4% of our total sample declare to be female and 29.4% male and 18.2% did not disclose gender. The sample was divided equally across treatment arms, with around 2,200 subjects for each message and treatment duration and twice as many for the control. The randomization leads to a balance of pre-treatment characteristics as presented in Table 3.1. Subjects' number of steps before the treatment implementation are balanced across the treatment. However, our sample is unbalanced in terms of number of connections per day. Our experiment includes observations from each French region and, as shown in Figure 3.A2 (Appendix 3.A), the

[10]Registry number AEARCTR-0010467 (link: https://www.socialscienceregistry.org/trials/10467)

proportion of users relative to the regional population is uniformly distributed.

Table 3.1: Summary statistics

	Peer/1W	Peer/3W	Self/1W	Self/3W	Sunk-cost/1W	Sunk-cost/3W	Control	N
Female	0.65	0.64	0.65	0.62	0.65	0.64	0.64	14,215
	(0.67)	(0.82)	(0.50)	(0.22)	(0.52)	(0.87)		
% of users who do not share their location	0.42	0.45	0.46	0.43	0.44	0.44	0.45	17,377
	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)		
Average number of steps at baseline	6,050	5,994	6,037	6,144	5,989	6,011	6,092	17,377
	(0.68)	(0.34)	(0.59)	(0.61)	(0.31)	(0.43)		
% of subjects who walk $>$ 10,000 steps	0.12	0.11	0.12	0.13	0.12	0.13	0.12	17,377
	(0.73)	(0.32)	(1.00)	(0.75)	(0.89)	(0.71)		
% of subjects who walk $<$ 5,000 steps	0.44	0.46	0.45	0.44	0.46	0.45	0.44	17,377
	(0.91)	(0.31)	(0.48)	(0.91)	(0.22)	(0.41)		
Average number of connections at baseline	3.35	3.30	3.28	3.34	3.26	3.46	3.36	17,377
	(0.94)	(0.60)	(0.45)	(0.88)	(0.36)	(0.37)		
Probability of opening the app at baseline	0.53	0.52	0.53	0.51	0.52	0.52	0.53	17,377
	(0.56)	(0.44)	(0.75)	(0.08)	(0.37)	(0.30)		
N	2,223	2,114	2,160	2,201	2,132	2,193	4,354	

Notes: The table reports the means of each variable across each treatment arm.

3.2.3 Treatments

The study benefits of pre-treatment observations for the whole sample, which are collected for two weeks to assess baseline levels. In these two weeks, all users received the basic reminder usually sent by the application. This is the same message that the control group will receive for all the duration of the experiment and that will be sent in the post-treatment period.

- **Basic reminder:** "Don't forget to validate your steps! It's time to increase your prize fund". We consider this message as a neutral reminder of the existence of the app and how it works.

After the baseline period, we randomly assigned the users to one of six treatment arms in a 3x2 design. A seventh group that serves as the control group receives the basic reminder for the entire duration of the study. In addition to the reminder variations, the other condition that varies orthogonally is the duration of the treatment. The structure and the content of the treatments is detailed as below.

1. **Sunk-cost:** "You have validated your steps over the last few days, do it again to win more *Wards*" The message for the sunk-cost group emphasizes the fact that the subject has already invested time and effort in walking and using the app. This should represent a push to continue on the path to be able to reach the benefits.

   (a) **Sunk-cost/1W:** This arm receives the sunk-cost treatment for one week.

   (b) **Sunk-cost/3W:** This arm receives the sunk-cost treatment for three weeks.

2. **Peer:** "Today, the WeWarders have walked XXX steps on average, what about you?" Since *WeWard* offers the possibility to connect with friends or strangers and it has many features able to create a sense of community, the peer comparison group receives a message that frames the social aspect of the app and aims at leveraging users' curiosity and competitive behavior. While the competitive feature can work differently according to the relative position, curiosity about social comparison should promote the attention to the number of steps.

   (a) **Peer/1W:** This arm receives the peer treatment for one week.

   (b) **Peer/3W:** This arm receives the peer treatment for three weeks.

3. **Self:** "Last week you walked an average of XXX steps, how many did you walk today?" The self-comparison group receives a message that highlights own past behavior and aims at leveraging users' self-improvement. This could work differently according to improvement possibilities.

   (a) **Self/1W:** This arm receives the self-comparison treatment for one week.

   (b) **Self/3W:** This arm receives the self-comparison treatment for three weeks.

As mentioned earlier, we collected users' data for the two weeks before the treatment implementation (pre-treatment), and for the two weeks after (post-treatment). During the pre and post-treatment, all subjects received the basic reminder.

3.3 Hypotheses

The focus of our study is the relationship between the information treatments administered to the app users and the average number of steps they walk. Since the app tracked the steps and is focused on walking, we study the engagement with the app as proxy of attention to the healthy habit.

Sending reminders is an effective and relatively inexpensive way of promoting physical exercise (Calzolari & Nardotto, 2017; Habla & Muller, 2021). However, generic messages have a limited impact (Costa & Kahn, 2013; Halpern, 2015; Peer *et al.*, 2020). Personalized messages have been found efficient to improve academic and driving performances (O'Connell & Lang, 2018; Choudhary *et al.*, 2022) or to ensure enrollment in college programs among disadvantaged populations (Castleman & Page, 2015). Their introduction is beneficial to induce new behaviors (e.g. entering college) and to enhance existing behaviors. Punctual messages can also promote exercising in peculiar samples (Golbus *et al.*, 2024). Relying on this evidence, we send specifically designed messages, and expect them to be effective in increasing the number of steps compared to the control group. Moreover, our messages add to the reminder emphasis on one of the behavioral aspects of the app therefore potentially amplifying the effect. In addition to the comparison with the control, a direct comparison between the different treatments allows us to determine which is the behavioral aspect most effective in impacting *WeWards* users' behavior.

Hypothesis 1: *Walking and engagement with the app increase when users receive specific messages leveraging behavioral features compared to standard reminders.*

Besides the content of the message, the duration of the treatment plays a role in affecting the primary outcome. In this study, we change the content of the reminders, and expose groups to different duration of these messages. In the health domain, the timing of the intervention can be relevant for its effectiveness (Head *et al.*, 2013; Steiner *et al.*, 2018;

Hardeman *et al.*, 2019). While variation of the duration and intensity of the physical activity promoted has been studied (Charness & Gneezy, 2009), our focus is on the duration of the messaging aimed to promote the behavior. We predict different effects in increasing in the number of steps for the long and short run treatments. Subjects in treatments Self/3W, Peer/3W, and Sunk-cost/3W benefit from a longer period of personalized reminders and that can imply more persuasiveness compared to those who receive the framed reminder for only one week. Moreover, we expect a similar effect on subjects' engagement with the app since a longer period of personalized behavioral messages can help build the habit of consulting the app.

Hypothesis 2: *Walking and engagement with the app increase when users receive messages for a longer period of time.*

Previous studies specifically target subjects that do not regularly exercise (Acland & Levy, 2015), or they analyze the treatment effect distinguishing between regular and non-regular (Charness & Gneezy, 2009) expecting that not regular can be affected the most by the intervention. In our sample, all subjects can be considered motivated, since they downloaded the app. However, gathering previous data from *WeWard*, we can distinguish between subjects who walk an average amount of steps or above, and subjects who walk little and therefore can be considered sedentary. Exploiting this distinction, we can determine the effect of the messages in changing the number of steps among regular walkers and non-regular. Since the potential for improvement is higher for non-regular (Charness & Gneezy, 2009), we expect the treatment effect to be greater for low-walkers.

Hypothesis 3: *The treatment is more effective among low-walkers.*

3.4 Empirical strategy

To measure the effects of the treatments on the attention and the walking habits of *WeWard*' users, we first compare the effect of being treated with any of the reminders on several outcome variables, and then we compare a specific outcome in each of our treatment groups with the same outcome variable among respondents in the control group that received the generic message for the entire duration of the intervention. More specifically, we estimate the following econometric model:

$$\begin{aligned} Y_{i,n} = \beta_0 + \beta_1 \mathit{Treated} + \beta_2 T_{i,control} \times \mathit{Treated} + \beta_3 T_{i,sunk} \times \mathit{Treated} \\ + \beta_4 T_{i,peer} \times \mathit{Treated} + \beta_5 T_{i,self} \times \mathit{Treated} + \epsilon_i \end{aligned} \tag{3.1}$$

where $Y_{i,n}$ is a set of n outcomes of interest: a dummy capturing whether the respondent opened the app on a given day, the number of connections to the app per day, the weekly average number of steps, and the daily number of steps. To study attention to steps we look at the engagement with the app, we define a dummy variable that is equal to 1 when the respondent opens the app on that day and 0 otherwise. Moreover, we assess the number of connections and the number of actions the user triggers on the app within a day. When we use this set of variables as the outcome in the regression, we keep the entire sample of observations. When analysing users' walking habits, the weekly average number of steps (*Weekly steps*) is calculated as the average number of steps per week and per individual, after restricting the sample to those individuals with at least one record per week. The daily number of steps (*Daily steps*) is the number of steps walked at any given day, by those individuals who opened the app every day over the intervention period.

Treated is a dummy equal to 1 after week 2 and before week 4 (for one week's treatments) or week 6 (for three weeks' treatments), thus indicating that we are considering the treatment period. As specified in each of our regressions, the treatment effect is estimated in comparison with the pre-treatment (*Pre-Treatment*) or the post-period (*Post-Treatment*). In addition, we assess the effect of new reminders between the pre-treatment period and the post-treatment

period.

$T_{control}$, T_{sunk}, T_{peer}, T_{self} are dummies equal to 1 if subject i was randomly assigned to each of the 3 treatments or the control group, receiving the generic reminder for the entire duration of the intervention. ϵ_i is the error term.

Hence, β_0 captures the average outcome in the control group in the two weeks before the start of the treatment. β_1 captures the evolution of the variable of interest after the implementation of the treatment compared to the two weeks of pre-treatment (or to the two/four weeks of post-treatment). Finally, β_2-β_3, β_2-β_4 and β_2-β_5 assess the difference between step variation (or engagement through the app) in the treatment groups throughout the experiment, compared with the control group.

3.5 Results

As we are interested in the effect of the treatments on both the attention to walking and the walking habits, our results in Subsection 3.5.1 focus on the app engagement, while in Subsection 3.5.2 focus on users' walking behavior. Given that opening the application is a necessary condition for *WeWard* to retrieve users' activity data, our sample is confronted with an attrition problem. We take this issue into account by running the analysis on walking habits on two different sub-samples in addition to the original one. In Subsection 3.5.2.1, we focus on the outcome Average number of steps per week, using the sub-sample of subjects who open the app at least one day in a week per the whole duration of the experiment. Then we restrict the sample to participants who always validate their steps during the intervention. Operating this restriction, we present the results on the daily number of steps, a more granular measure compared to the weekly average, focusing on the most committed users.

3.5.1 Engagement with the app

Even if downloading the app is a signal of interest in step tracking and walking, subjects' app behavior is various. Only 3.62% of our initial sample have opened the app every day. Application-related behaviors inform us about users' attention and motivation, which is essential to understand before moving to the study of walking habits. Indeed, targeting engagement with the app could improve walking behaviors. The number of steps is positively correlated with the number of times the subject opens and interacts with the app (resp. Pearson's r = 0.27 and r = 0.26, $p = 0.00$). This interrelation between the physical activity and the attention on its measurement can provoke a virtuous cycle if positively activated. We start analyzing the determinants of opening the app, in subsection 3.5.1.1.[11] Then, we focus on the number of connections, a variable that we use as a proxy for the degree of interaction with the app.

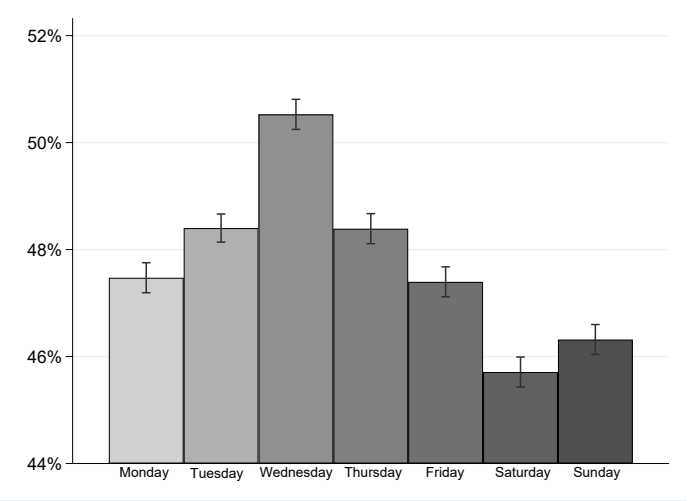
3.5.1.1 Open the app

On average, subjects open the application 4.37 times per day over the 50 days of the experiment. However, subjects' use of the app is very heterogeneous. Half of the users did not open the app for 32 days and, on average, users do not open the app for 3.48 days in a row. 1,143 individuals opened *We Ward* on only once during the period of the experiment.

Figure 3.2 presents the share of users who open *We Ward*, based on the day of the week. Users open significantly less the app during the week-end (χ^2, $p = 0.00$). They also tend to open slightly more the application during Sunday compared to Saturday (χ^2, $p = 0.00$). Figure 3.A3 in the Appendix shows that the attention to the steps differs also by gender. Each day of the week females are more likely to open the app, compared to men (χ^2, $p = 0.00$).

[11] Our sample is composed by all subjects (N = 17,377) who were active on the app at least once. We remove from the analysis users who never opened the application (N = 2,212) and those who have a null amount of steps for the 50 days of the experiment (N = 598).

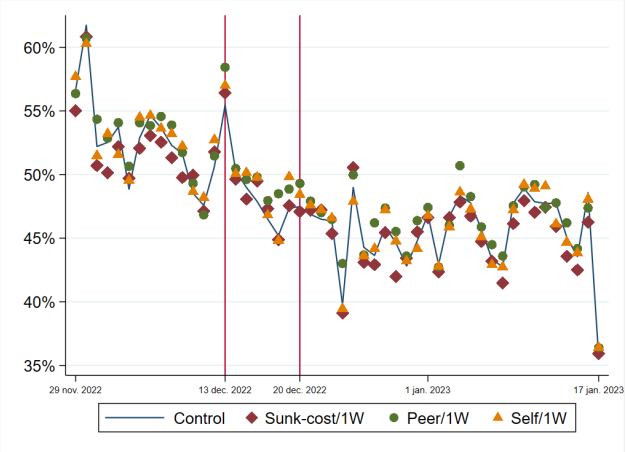
Figure 3.2: Share of our users who open *WeWard*, by day of the week.



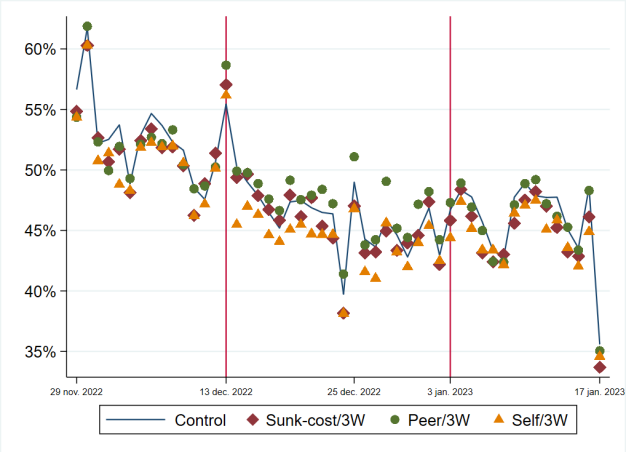
Figures 3.3a and 3.3b illustrate the evolution of the proportion of users who open the application during the different periods of observation, by treatment group and duration of the treatment. The two red lines represent respectively the beginning and the end of the treatment period. The figure shows that the proportion is not stable over time. It presents short term variation with weekly negative peaks, due to the fact that users open less *WeWard* during the weekends and a downward trend, attributable to the fact that the experiment took place before and during the Christmas holidays. Indeed, users open significantly less the application on Christmas Eve (24th of December) and New Year's Eve (31st of December). This effect also occurs on the 25th of December and the 1st of January (χ^2, $p = 0.00$ and $p = 0.03$ for New Year). We do not observe a clear effect of treatments in Figure 3.3a while the impact of treatments appears more nuanced in Figure 3.3b. The negative trend in the probability of opening the app in the three weeks' treatments seems to be attenuated by the treatment Peer/3W compared to the other treatments.

Figure 3.3: Time trend of share of subjects who open *We Ward*

(a) One-week treatments



(b) Three-weeks treatments



Tables 3.2-3.B3 report the effects of the treatments on the probability of opening the app. First, we present the results comparing the treatment period versus the pre-treatment period (Table 3.2), to highlight any treatment effect. Second, we present the results comparing the post-treatment period versus the treatment period (Table 3.3), to test for the presence of a rebound effect once the treatment is removed. Finally, we test for long-run pre-post treatment effects by comparing the post-treatment and pre-treatment periods (Table 3.B3). The Tables present six specifications, using a Logit model with individual fixed effects. Columns 1-3-5 focus on the one week's treatments, while columns 2-4-6 on the three weeks' treatments. In columns 1-2, we run the basic specifications with no controls. In columns 3-4 we add time effect controlling for the day of the week. In columns 5-6, we add a time trend to take into account the fact that we run the experiment before and during the Christmas's holiday.

In Table 3.2, the treatment effect is captured by comparing the treatment period with the pre-treatment period. We observe a negative and significant coefficient of *Treated* in columns 1 to 4, indicating that subjects significantly open less the application in the treatment period compared to the pre-treatment period. The coefficients range from -0.27 for the one-week treatments to -0.43 for the three-weeks treatments. However, when we include the time trend (columns 5-6), the coefficients of *Treated* change their sign (to 0.18, and 0.02). This confirms

Table 3.2: Effects on the probability of opening the app: **Treatment** vs. **Pre-treatment**

	Probability of opening the app					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.27***		-0.27***		0.18***	
	(-14.50)		(-14.55)		(7.15)	
Peer/1W × Treated	0.08**		0.08**		0.08**	
	(2.61)		(2.62)		(2.60)	
Self/1W × Treated	0.05		0.05		0.05	
	(1.42)		(1.43)		(1.42)	
Sunk-cost/1W × Treated	0.07*		0.07*		0.07*	
	(2.05)		(2.06)		(2.06)	
Treated		-0.43***		-0.43***		0.02
		(-31.45)		(-31.50)		(0.93)
Peer/3W × Treated		0.13***		0.13***		0.13***
		(5.58)		(5.59)		(5.60)
Self/3W × Treated		-0.01		-0.01		-0.01
		(-0.24)		(-0.24)		(-0.25)
Sunk-cost/3W × Treated		0.04		0.04		0.04
		(1.64)		(1.64)		(1.64)
Observations	195,468	352,240	195,468	352,240	195,468	352,240
Number of subjects	9,308	10,064	9,308	10,064	9,308	10,064
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: Logit model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

the fact that our results are affected by the downward trend. When we look at the marginal effect of receiving the framed reminders, we see that receiving the treatment that highlights the peer comparison, for either 1 or 3 weeks, significantly increases the probability of opening the app. This treatment is indeed able to mitigate the downward trend by 0.08 if it is received for one week, and by 0.13 if it is received for 3 weeks. We observe a positive, while less significant, effect of the treatment Sunk-cost/1W, stable across the different specifications.

In Table 3.3, we compare the post-treatment period with the treatment period, to investigate whether the effect holds once the reminder is removed. As for Table 3.2, we observe

Table 3.3: Effects on the probability of opening the app: **Post-treatment** vs. **Treatment**

	Probability of opening the app					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-0.24***		-0.24***		-0.16***	
	(-13.83)		(-13.90)		(-8.02)	
Peer/1W × Treated	-0.07*		-0.07*		-0.07*	
	(-2.35)		(-2.36)		(-2.35)	
Self/1W × Treated	-0.07*		-0.07*		-0.07*	
	(-2.26)		(-2.27)		(-2.27)	
Sunk-cost/1W × Treated	-0.07*		-0.07*		-0.07*	
	(-2.22)		(-2.22)		(-2.22)	
Treated × Post		-0.04**		-0.04**		0.38***
		(-2.74)		(-2.84)		(19.15)
Peer/3W × Treated		-0.11***		-0.11***		-0.11***
		(-4.59)		(-4.59)		(-4.60)
Self/3W × Treated		0.00		0.00		0.00
		(0.09)		(0.09)		(0.10)
Sunk-cost/3W × Treated		-0.07**		-0.07**		-0.07**
		(-2.79)		(-2.79)		(-2.79)
Observations	351,504	350,964	351,504	350,964	351,504	350,964
Number of subjects	9,764	9,749	9,764	9,749	9,764	9,749
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: Logit model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

that the negative coefficients of the variable *Treated* in columns 1 to 4 are mitigated by the introduction of the time trend in columns 5-6. When we look at the specific effect of the framed reminders, we are particularly interested in whether the positive effect of the Peer treatment holds once subjects do not receive the reminder anymore. The coefficients are negative and significant, especially for the groups exposed to the reminder for 3 weeks (-0.11). Moreover, the magnitude is comparable to the positive effect observed in Table 3.2: 0.13 vs 0.11. This confirms previous evidence on the use of reminders and notifications: they seem to work while they are in place, while in the long-run (or at least once the tool is removed)

subjects tend to return to their previous behavior, or in extreme cases to set on even worse levels.

To assess if this is the case, we compare the pre and post-treatment periods in Table 3.B3, in the Appendix. None of the coefficients on the customized reminders is significant, neither for the one-week groups nor for the three-weeks groups. The combined evidence from Tables 3.2 and Table 3.B3 indicates that the effect of messages leveraging the comparison with peers does not persist when they are no longer implemented. Our results add on evidence that customized reminders, and especially the ones leading to social or self comparison, are indeed efficient only when implemented (Slaunwhite *et al.*, 2009).

After presenting the effect of the reminders on the probability of opening the app as first outcome to measure the effect that the messages had on the attention on the step tracking, in the next section we investigate whether they have an impact on the quantity of interaction with the app, before moving to the results on walking habits. Together, these two outcomes reflect the modification in the attention on the topic unleashed by our messages.

3.5.1.2 Activity on the app

In this section, the aim is to go beyond opening the app and investigate whether the reminders influence how much the users interact with the app. To measure this, we use the daily number of connections to the application.

Tables 3.4-3.5 (and Table 3.B4 in the Appendix) present the results of the effect of the messages on the number of connections, comparing the pre-treatment, treatment and post-treatment periods. We use an OLS model with individual fixed effects. From Table 3.4 we see that, as for the probability of opening the app, introducing the time trend (columns 5-6) reduces the negative coefficient of the variable *Treated*. When we look at the marginal contribution of the customized messages on the overall effect we report evidence that the number of connections increases for subjects exposed for three weeks to the message that

leverages self-comparison (0.17). This effect is stable across different specifications (columns 2-4-6). We also observe a mildly significant effect of the Sunk-cost Treatment, in the three weeks group. Overall, being exposed for a longer period to the reminders seem to affect more effectively the engagement with the app in terms of intensity of activity.

When we compare the post-treatment to the treatment period (Table 3.5), we can confirm that the coefficients for the messages that were effective while in place, are now negative and higher in magnitude (-0.26 for the Self-comparison 3W treatment, and -0.16 for the Sunk-cost/3W treatment). Indeed, in line with the results we obtained on the probability of opening the app, we do not find any persistent effect of the intervention once the customized messages are removed. Table 3.B4 in the Appendix shows the results of the comparison between post and pre-treatment and, as expected, none of the coefficients is statistically significant.

Table 3.4: Effects on *Number of connections*: **Treatment** vs. **Pre-treatment**

	Number of connections					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.32***		-0.32***		-0.13*	
	(-7.94)		(-8.07)		(-2.41)	
Peer/1W × Treated	0.00		0.01		0.01	
	(0.07)		(0.10)		(0.08)	
Self/1W × Treated	0.07		0.07		0.07	
	(1.07)		(1.04)		(1.06)	
Sunk-cost/1W × Treated	0.06		0.06		0.06	
	(0.90)		(0.86)		(0.90)	
Treated		-0.40***		-0.40***		-0.16***
		(-12.81)		(-12.92)		(-3.60)
Peer/3W × Treated		0.01		0.01		0.00
		(0.10)		(0.10)		(0.09)
Self/3W × Treated		0.17**		0.17**		0.17**
		(3.05)		(3.06)		(3.05)
Sunk-cost/3W × Treated		0.11*		0.11*		0.11*
		(2.03)		(2.01)		(2.00)
Constant	3.94***	4.14***	4.03***	4.19***	422.87***	323.75***
	(281.74)	(283.34)	(126.59)	(151.56)	(5.50)	(7.75)
Observations	110,034	170,101	110,034	170,101	110,034	170,101
Number of subjects	10,119	10,583	10,119	10,583	10,119	10,583
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.5: Effects on *Number of connections*: **Post-treatment** vs. **Treatment**

	Number of connections					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-0.05		-0.05		-0.12**	
	(-1.31)		(-1.42)		(-2.85)	
Peer/1W × Treated	-0.06		-0.06		-0.06	
	(-0.92)		(-0.92)		(-0.91)	
Self/1W × Treated	-0.01		-0.01		-0.01	
	(-0.13)		(-0.12)		(-0.13)	
Sunk-cost/1W × Treated	0.03		0.03		0.03	
	(0.47)		(0.48)		(0.47)	
Treated × Post		0.13***		0.13***		0.27***
		(4.20)		(4.06)		(5.78)
Peer/3W × Treated		-0.10		-0.10		-0.10
		(-1.75)		(-1.74)		(-1.75)
Self/3W × Treated		-0.26***		-0.26***		-0.26***
		(-4.72)		(-4.69)		(-4.73)
Sunk-cost/3W × Treated		-0.16**		-0.16**		-0.16**
		(-2.90)		(-2.86)		(-2.90)
Constant	3.81***	3.89***	3.79***	3.89***	-90.55**	172.03***
	(188.71)	(301.95)	(124.86)	(141.69)	(-3.17)	(4.08)
Observations	167,191	164,857	167,191	164,857	167,191	164,857
Number of subjects	10,106	10,081	10,106	10,081	10,106	10,081
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

3.5.2 Walking habits

In the following sections, we focus on walking behaviors and the effect of our treatments in triggering virtuous behaviors. We start analyzing the walking habits of the full sample of users. To test the robustness of our findings, we then focus on the weekly average number of steps, and finally on strongly committed users who open the app every day. Table 3.6 reports the effect of our reminders on the number of steps walked, comparing the treatment period with the baseline. We use an OLS model with individual fixed effects. As one can see, we observe that being treated for both one week and three weeks affects the number of steps that subjects walk in a day. After adjusting for the time trend, the effect is positive and ranges from 315.30 (for the 1 week treatments) to 357.64 (for the 3 weeks treatments). None of the customized reminders distinguish itself in changing the individual number of steps walked. As we do not find any impact of the different messages, in the main text, we only report the results of the comparison between the treatment period and the baseline. For the comparison of the post-treatment and treatment, and post-treatment and pre-treatment refer to Tables 3.B5-3.B6 in the Appendix.

3.5.2.1 Weekly and daily steps

To exploit the temporal dimension of our data, we perform panel analyses. We restrict the sample to participants who recorded at least one observation per week, in each week of the experiment. This procedure leads to a final sample of 8,941 subjects, and allows us to conduct a balanced panel analysis with *customerID* as the panel variable and *week* as the time variable. Our variable of interest is the variable that reports the individual average number of steps per week, which we defined as *Weekly Steps*.

Figures 3.4a and 3.4b illustrate the evolution of the variable *Weekly Steps* along the experiment, by treatment group. As shown in the graphics, the average number of steps per week decreases for all the groups, with a large drop in weeks 4 and 5, and then increases.

Table 3.6: Effects on *Steps walked*: **Treatment** vs. **Pre-treatment**

	Steps walked					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-224.44***		-235.70***		315.30***	
	(-5.58)		(-5.95)		(5.91)	
Peer/1W × Treated	-10.34		0.36		-8.86	
	(-0.15)		(0.01)		(-0.13)	
Self/1W × Treated	63.73		54.40		61.80	
	(0.92)		(0.79)		(0.89)	
Sunk-cost/1W × Treated	-59.87		-68.68		-61.43	
	(-0.85)		(-0.99)		(-0.87)	
Treated		-648.96***		-653.92***		357.64***
		(-20.85)		(-21.30)		(8.12)
Peer/3W × Treated		-35.36		-39.62		-37.76
		(-0.65)		(-0.74)		(-0.70)
Self/3W × Treated		60.59		55.14		60.45
		(1.11)		(1.02)		(1.11)
Sunk-cost/3W × Treated		11.98		0.29		6.31
		(0.22)		(0.01)		(0.12)
Constant	6274.28***	6348.75***	6370.29***	6347.74***	1.19e+06***	1.36e+06***
	(447.11)	(431.67)	(202.22)	(231.16)	(15.49)	(32.36)
Observations	109,794	169,746	109,794	169,746	109,794	169,746
Number of subjects	10,121	10,585	10,121	10,585	10,121	10,585
Day of the week			Yes	Yes		
Time trend					Yes	Yes

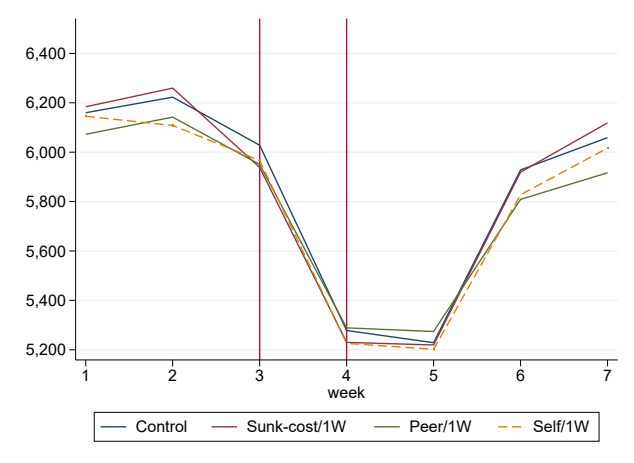
Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Since the fourth week starts on the 20th of December and the fifth week ends on the 2nd of January, this decline is mainly due to a seasonal effect (i.e. the Christmas holidays). This effect seems smaller for the Self/3W treatment, while the decline appears larger for Peer/3W. For those who receive the treatment for one week, the difference between treatments is not apparent.

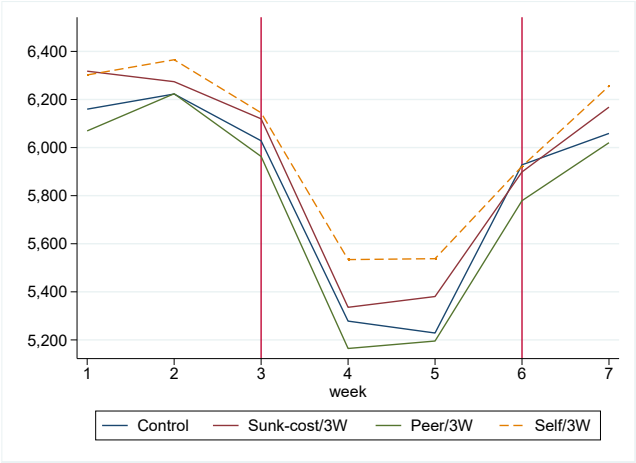
We further assess the effect of different messages and treatment's duration on walking behavior, using an OLS model. Table 3.7 presents the results with *Weekly Steps* as dependent variable. The table follows the same structure of previous tables. In columns 1-3-5, we report the results for the one week's treatments, and in columns 2-4-6 we report the results for the

Figure 3.4: Average number of steps per week by treatment group.

(a) One week treatments



(b) Three weeks treatments



three weeks' treatments. In specifications 3-4, we control for the day of the week, while in 5-6 we add a time trend. As shown by the significant and negative coefficient of *Treated*, subjects walk less per week on average after the implementation of the treatment. Overall, the treatments do not have an impact on the average number of steps per week. The results are in line with the results presented in Table 3.7, confirming that customized messages are not effective in increasing the number of steps, even when we restrict the sample to users that open the app at least once per week. We only report the results of the comparison between the treatment period and the baseline. For the comparison of the post-treatment and treatment, and post-treatment and pre-treatment refer to Tables 3.B7-3.B8 in the Appendix.

To test the results that we find on the weekly average of the number of steps and on the amount of steps walked, we further restrict our sample and exclude all subjects who have not opened the application (and then allow the collection of their steps) every day, i.e. for 50 days. This procedure significantly reduces the sample size leading to a number of subjects equals to 553 but allows us to conduct a balanced panel analysis with *customerID* as the panel variable and *date* as the time variable. Since this new dataset has daily data with no missing day, we can study the number of steps per day, *Daily Steps*, rather than per week, *Weekly Steps*. Furthermore, this method enables us to explore the heterogeneity of walking

Table 3.7: Effects on *Weekly Steps*: **Treatment** vs. **Pre-treatment**

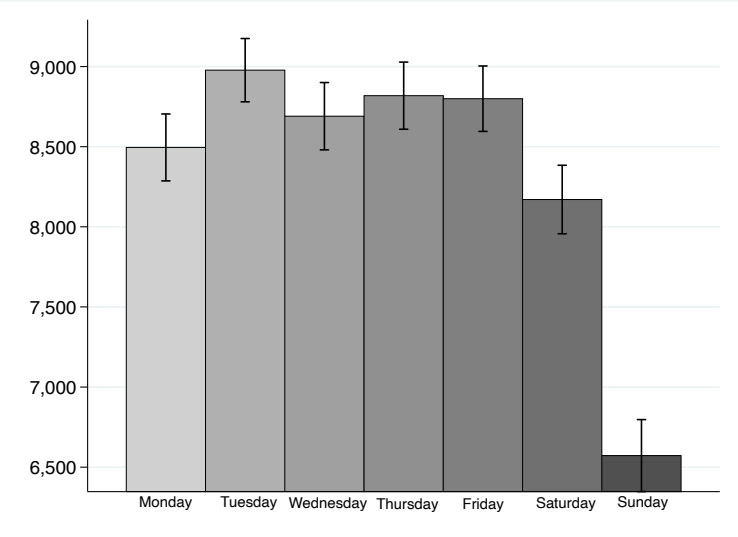
	Weekly steps					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-162.99***		-163.24***		-217.97***	
	(-3.32)		(-3.33)		(-3.35)	
Peer/1W × Treated	6.33		6.80		5.73	
	(0.07)		(0.08)		(0.06)	
Self/1W × Treated	2.75		2.99		2.03	
	(0.03)		(0.04)		(0.03)	
Sunk-cost/1W × Treated	-120.09		-118.82		-120.05	
	(-1.32)		(-1.30)		(-1.32)	
Treated		-679.52***		-678.69***		-32.42
		(-15.58)		(-15.56)		(-0.57)
Peer/3W × Treated		-26.18		-26.09		-26.89
		(-0.35)		(-0.35)		(-0.36)
Self/3W × Treated		84.76		80.59		80.49
		(1.04)		(0.99)		(0.98)
Sunk-cost/3W × Treated		-4.52		-3.16		-4.06
		(-0.06)		(-0.04)		(-0.05)
Constant	6166.54***	6230.98***	6180.98***	6245.63***	-1.16e+05	857920.44***
	(578.93)	(363.63)	(136.53)	(163.95)	(-1.15)	(14.98)
Observations	16,914	27,770	16,914	27,770	16,914	27,770
Number of subjects	5,638	5,554	5,638	5,554	5,638	5,554
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

practices within the week. Figure 3.5 illustrates the mean of *Daily Steps* by day of the week. Overall, subjects walk more during week days than during the week-end.

Figures 3.6a and 3.6b present the evolution of *Daily Steps* across the period of the experiment and offer more visibility on the determinants of the decrease in the number of steps, compared to Figures 3.4a and 3.4b. Indeed we find that the drop observed in the fourth week is mainly due to the 25th of December (Christmas Day) and that the same effect occurred in the fifth week and, more specifically, on the 1st of January (the day after New Year's Eve). The impact of the new reminders on the daily steps is not salient, except for users in

Figure 3.5: Average number of steps, by day of the week.

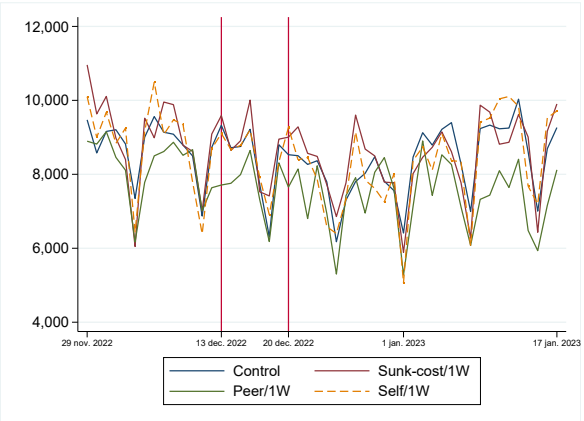


Notes: Confidence intervals at 95%.

Peer/1W who seem to experience steeper drops than the other treatments.

Figure 3.6: *Daily Steps*, by treatment group.

(a) One week treatments



(b) Three weeks treatments

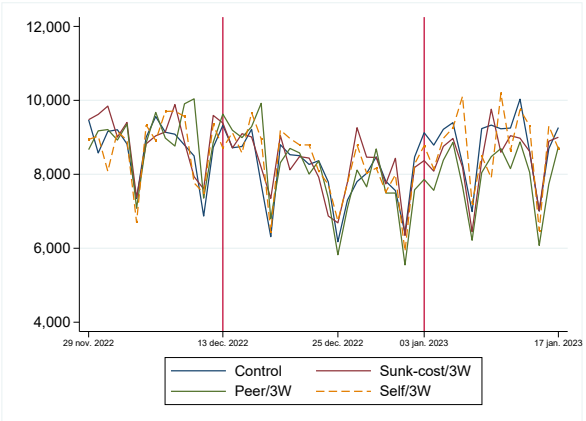


Table 3.8 presents the results of the regression ran on the sub-sample of users for whom we have daily data. Columns 1-3-5 present the results for the one week treatments, and columns 2-4-6 for the three weeks treatments. As for the *Weekly steps*, also here we do not find any significant effect of the customized reminders on the number of steps walked per day by the

most active users.[12] In the next section, we further test for the presence of heterogeneous effects by the users' level of activity.

Table 3.8: Effects on *Daily Steps*: **Treatment** vs. **Pre-treatment**

	Daily steps					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-324.49		-324.49*		684.15**	
	(-1.94)		(-1.97)		(3.01)	
Peer/1W × Treated	-201.64		-201.64		-201.64	
	(-0.67)		(-0.68)		(-0.67)	
Self/1W × Treated	-95.30		-95.30		-95.30	
	(-0.32)		(-0.33)		(-0.32)	
Sunk-cost/1W × Treated	46.94		46.94		46.94	
	(0.16)		(0.16)		(0.16)	
Treated		-715.99***		-715.99***		469.22*
		(-5.59)		(-5.66)		(2.56)
Peer/3W × Treated		-172.53		-172.53		-172.53
		(-0.75)		(-0.76)		(-0.75)
Self/3W × Treated		156.21		156.21		156.21
		(0.67)		(0.68)		(0.68)
Sunk-cost/3W × Treated		-66.74		-66.74		-66.74
		(-0.31)		(-0.31)		(-0.31)
Constant	8706.62***	8826.22***	8832.03***	8952.41***	2.22e+06***	1.57e+06***
	(138.23)	(138.66)	(63.91)	(77.26)	(6.56)	(9.06)
Observations	7,224	12,425	7,224	12,425	7,224	12,425
Number of subjects	344	355	344	355	344	355
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

3.5.2.2 Heterogeneity

In the following, we investigate heterogeneous effects according to subjects' pre-treatment activity levels, following the approach of Charness & Gneezy (2009). We define four categories of walkers based on the quartiles of the distribution of steps walked: inactive, somehow active, active, very active. We replicate the exercise using first the restricted sub-sample of users

[12] Tables 3.B9-3.B10 in the Appendix report the results of the comparison between post-treatment and treatment period, and post-treatment and pre-treatment period.

who opened the app every day during the experiment, and then the sub-sample of users for whom we have at least one observation per week. For the sub-sample of strongly committed users, *inactive* are users who walk less than 5,445 steps during the pre-treatment period, *somehow active* users walk between 5,445 and 7,510 steps, *active* users walk between 7,510 and 10,765 steps, and *very active* users walk more than 10,765 steps. For the sub-sample of users for whom we have at least one observation per week, *inactive* are users who walk less than 3,800 steps during the pre-treatment period, *somehow active* users walk between 3,800 and 5,570, *active* users walk between 5,570 and 7,710 steps, and *very active* users walk more than 7,710 steps.

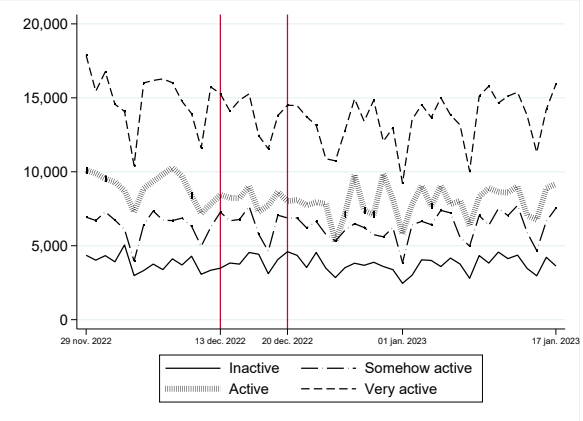
Strongly committed users walk on average more than those who did not open the application every day. Figures 3.7a and 3.7b illustrate the evolution of the average *Daily Steps*, by category while Figures 3.8a and 3.8b focus on *Weekly Steps*, by category. Figure 3.7a represents the evolution of *Daily Steps* for those who receive the treatment for one week, and Figure 3.7b for three weeks'. The trends shown in Figures 3.7a and 3.7b appear less stable over time than the ones shown in Figures 3.8a and 3.8b, because daily data reflect the fluctuations in steps between week days and weekend days. Also, they provide additional information on the Christmas' effect which is mainly driven by very active users. The drop on the 25th of December seems higher for *very active* and *active* users, compared to *somehow active* and *inactive* users. Furthermore, fluctuations in general appear more weekends dependent for *very active* and *active* users.

In the sub-sample of users who opened the app at least once per week throughout the experiment, the average number of steps per day is 5,848.[13] If we look at Figures 3.8a and 3.8b, the decline in the number of steps during the treatment period (observed in Figures 3.4a and 3.4b) appears to be mainly due to *somehow active*, *active*, and *very active* users. While their trend tend to diminish during weeks 4 and 5, *inactive* users' trend is relatively

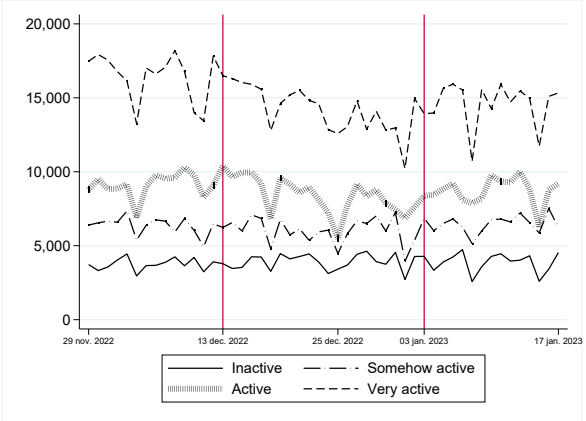
[13]According to the WHO, walking less than 5,000 steps per day is considered sedentary. Therefore, it is worth specifying that our categorization is based on the data from our sample, in which a relatively small number of users (i.e. 1,537; which represents 17% in the pre-treatment period) walk in compliance with the WHO's recommendations (i.e. 10,000 per day).

Figure 3.7: *Daily Steps*, by category.

(a) One week treatments



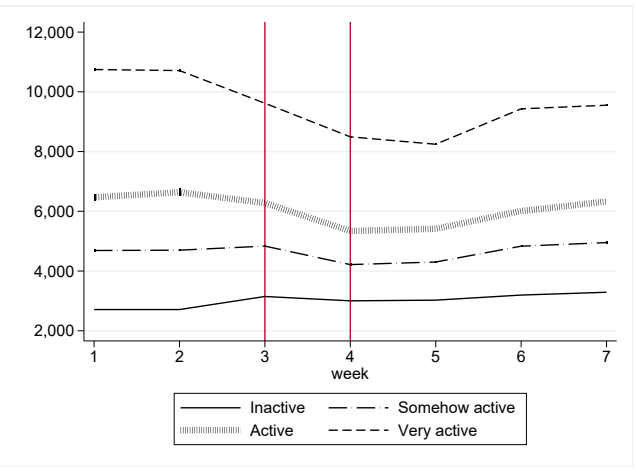
(b) Three weeks treatments



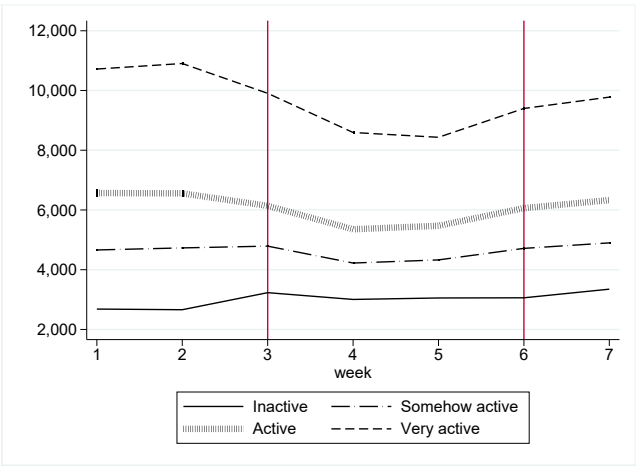
stable. Users who walk the least seem to experience, both in one week's and three weeks' treatments, a slight increase in their average number of steps during week 3. Figure 3.A4 in the Appendix demonstrates that the decrease of *Weekly Steps* for *very active* users is indeed significant while *inactive* users experience a significant increase (Mann-Whitney, $p = 0.00$).

Figure 3.8: *Weekly Steps*, by category.

(a) One week treatments



(b) Three weeks treatments



In order to assess whether our treatments have a different impact on the different categories of walkers, we split the sample and run the regression with *Weekly Steps* as dependent variable. Table 3.9 report the results of the comparison between the treatment period and

pre-treatment period. Columns 1-3-5-7 show the specifications for the one week treatments, while columns 2-4-6-8 for the three weeks treatments. Columns 1-2 report the results for *inactive* users, columns 3-4 for *somehow active* users, columns 5-6 for *active* users, and columns 7-8 for *very active* users. As shown by the significant and positive coefficient of *Treated* in specifications 1-2, the treatment period has a significant positive effect on *Weekly Steps* of *inactive* users. For less active users, we observe the opposite effect since the coefficient of *Treated* is negative in columns 5-8. The magnitude of the coefficients increase from columns 5-6 to 7-8, showing that the more active the user is the more is negatively affected by the treatment. Overall, customized messages do not have an impact on the number of steps of any categories of users. The positive effect on *inactive* users that we find comparing the treatment period with the pre-treatment period, is still there when we compare the post-treatment period with the pre. See Table 3.B12 in the Appendix. This suggests evidence that receiving reminders with behavioral features through app notifications might be effective in increasing walking also in the longer run but only for *inactive* users.

Table 3.9: Effects on *Weekly steps* by initial level of walking: **Treatment** vs. **Pre-treatment**

	Weekly steps							
	Inactive		Somehow active		Active		Very active	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	643.03***		27.79		-290.49**		-935.17***	
	(7.51)		(0.35)		(-2.98)		(-6.70)	
Peer/1W × Treated	-241.79		271.10		48.60		-74.24	
	(-1.80)		(1.95)		(0.30)		(-0.29)	
Self/1W × Treated	-205.66		48.15		260.42		-134.90	
	(-1.60)		(0.37)		(1.64)		(-0.68)	
Sunk-cost/1W × Treated	-246.13		107.00		-71.95		-245.27	
	(-1.85)		(0.68)		(-0.46)		(-1.02)	
Treated		308.85***		-382.28***		-1133.57***		-2392.63***
		(3.94)		(-4.42)		(-11.49)		(-16.77)
Peer/3W × Treated		-6.90		-37.05		-150.82		138.14
		(-0.06)		(-0.33)		(-1.26)		(0.70)
Self/3W × Treated		65.22		154.91		246.70		35.22
		(0.58)		(1.28)		(1.66)		(0.18)
Sunk-cost/3W × Treated		-74.43		108.72		-16.45		47.32
		(-0.70)		(0.93)		(-0.13)		(0.23)
Constant	2731.69***	2716.12***	4660.27***	4644.63***	6474.10***	6532.54***	10781.90***	10764.42***
	(107.65)	(87.03)	(142.36)	(124.63)	(170.10)	(146.32)	(185.02)	(164.37)
Observations	4,218	6,625	4,302	7,175	4,218	6,970	4,176	7,000
Number of subjects	1,406	1,325	1,434	1,435	1,406	1,394	1,392	1,400

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5-7 : one-week arms. Columns 2-4-6-8 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*. All specifications include weekly dummies.

3.6 Conclusions

We conduct a large-scale field experiment in collaboration with the step-tracking application *We Ward*. Our main objective is to test the effect of personalized reminders on engagement with the application (such as the probability of opening *We Ward*) and on walking behaviors. Specifically, we examine the evolution of several outcomes after the implementation of messages leveraging different behavioral features, as self-comparison, peer-comparison, and sunk-cost. Additionally, we assess the effect of treatments' duration on users' behaviors. Our results demonstrate the effectiveness of some type of reminders in increasing the engagement with the app. In particular, informing users about their peer performance induce them to open the app, while informing them about their own past achievements induce them to be more active on the app during the day. While we find that customized reminders are useful in drawing the attention to the app, they have little or no effect on the amount of steps

walked. Indeed, steps and walking behavior are harder to change and are strongly linked to subject's and external characteristics. Subjects are more sedentary during weekends and holidays suggesting that walking is somehow bundled to work activities, as intended more as a transportation mean that as a recreational health activity.

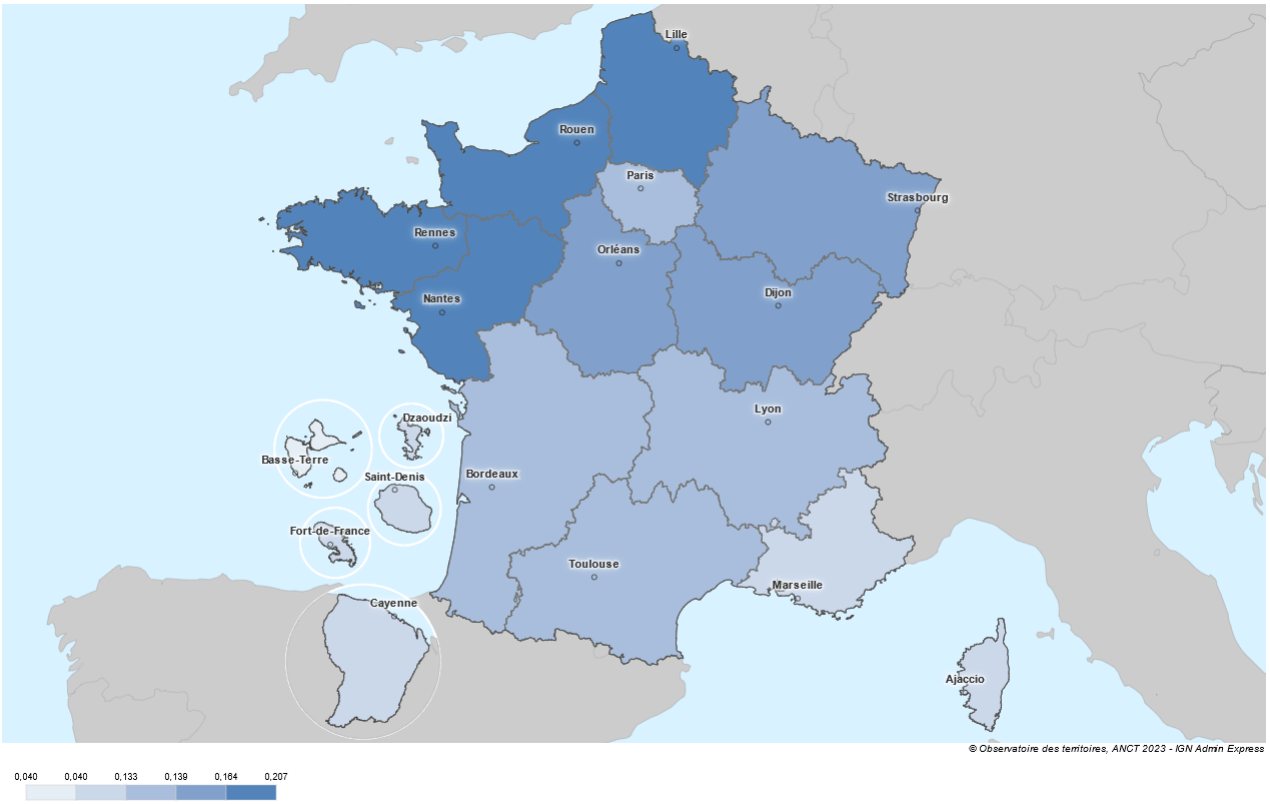
Contributing to the behavioral economics literature on the use of behavioral tools to address people's behavior, we explore the effects of a well-studied tool as framed messages are in a large scale experiment that allows us to distinguish between two outcomes: engagement and underlying behavior. Our findings highlight the importance of being able to measure actual behavior especially in the health context, and suggest that health behavior is hard to shift. Indeed, measuring health behaviors using app usage as a proxy could lead to overestimating the real effect of similar interventions.

3.A Additional Figures

Figure 3.A1: Screenshot of the app

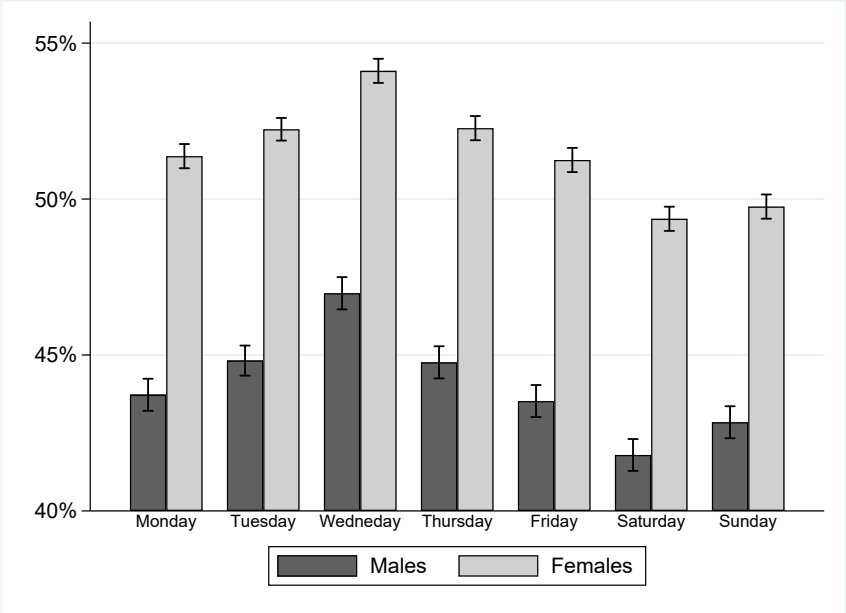


Figure 3.A2: Fraction of users in the population, by French region



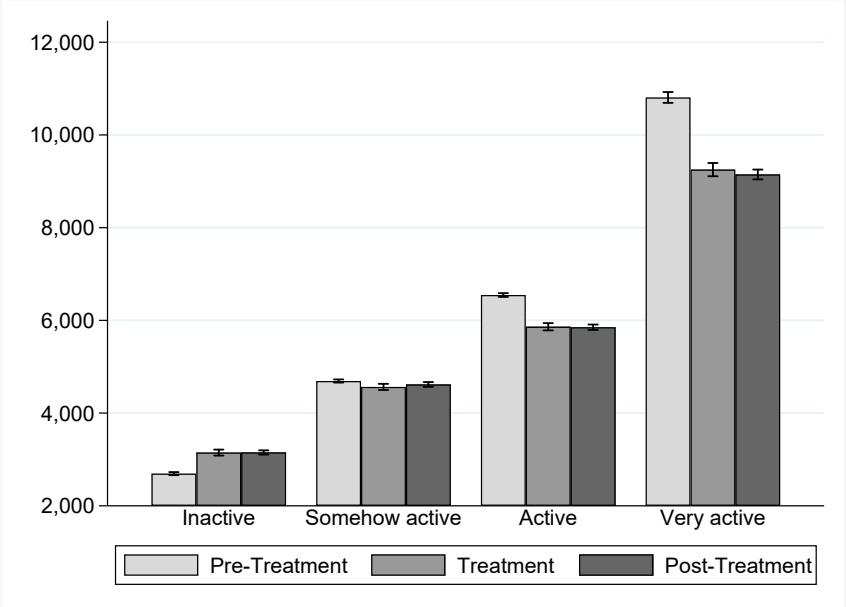
Notes: Figure 3.A2 illustrates the fraction of users (in ‰) in the population by region. Overall, the user rate is homogeneous. The lowest user rate is in DROM-COM (i.e. 0.04‰) and the largest is in Hauts-de-France (i.e. 0.21‰).

Figure 3.A3: Proportion of our users' sample who open *We Ward*, by gender and the day of the week



Notes: Number of subjects equal to 14,215. Note that, 3,162 users of our sample do not provide their gender through the app. Confidence intervals at 95%.

Figure 3.A4: *Weekly Steps* across the experiment window, by the category of walkers.



Notes: Confidence intervals at 95%.

3.B Additional Tables

Table 3.B1: Summary statistics

	Peer/1W	Peer/3W	Self/1W	Self/3W	Sunk-cost/1W	Sunk-cost/3W	Control	N
Female	0.65	0.64	0.65	0.62	0.65	0.64	0.64	14,215
Average number of steps at baseline	6,050	5,994	6,037	6,144	5,989	6,011	6,092	17,377
% of subjects who walk > 10,000 steps	0.12	0.11	0.12	0.13	0.12	0.13	0.12	17,377
% of subjects who walk < 5,000 steps	0.44	0.46	0.45	0.44	0.46	0.45	0.44	17,377
Average number of connections at baseline	3.35	3.30	3.28	3.34	3.26	3.46	3.36	17,377
Probability of opening the app at baseline	0.53	0.52	0.53	0.51	0.52	0.52	0.53	17,377
N	2,223	2,114	2,160	2,201	2,132	2,193	4,354	

Table 3.B2: Summary statistics: Strongly committed

	Peer/1W	Peer/3W	Self/1W	Self 3W	Sunk-cost/1W	Sunk-cost/3W	Control	N
Female	0.71	0.72	0.73	0.56	0.63	0.68	0.67	495
	(0.56)	(0.52)	(0.38)	(0.17)	(0.57)	(0.87)		
Average number of steps at baseline	8,228	8,918	8,837	8,762	8,998	8,972	8,732	553
	(0.53)	(0.83)	(0.90)	(0.97)	(0.75)	(0.77)		
% of subjects who walk > 10,000 steps	0.29	0.35	0.33	0.30	0.35	0.25	0.28	553
	(0.92)	(0.29)	(0.48)	(0.81)	(0.29)	(0.62)		
% of subjects who walk < 5,000 steps	0.23	0.20	0.21	0.30	0.18	0.11	0.26	553
	(0.61)	(0.35)	(0.42)	(0.59)	(0.23)	(0.01)		
Average number of connections at baseline	12.86	14.59	12.50	14.26	11.55	15.47	14.32	553
	(0.53)	(0.93)	(0.44)	(0.98)	(0.22)	(0.61)		
N	66	65	67	64	65	80	146	

Table 3.B3: Effects on the probability of opening the app: **Post-treatment** vs. **Pre-treatment**

	Probability of opening the app					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-0.49***		-0.49***		-0.31***	
	(-37.91)		(-37.96)		(-15.71)	
Peer/1W × Treated	0.02		0.02		0.02	
	(0.82)		(0.82)		(0.82)	
Self/1W × Treated	-0.02		-0.02		-0.02	
	(-0.81)		(-0.81)		(-0.81)	
Sunk-cost/1W × Treated	0.00		0.00		0.00	
	(0.12)		(0.12)		(0.11)	
Treated × Post		-0.47***		-0.47***		0.50***
		(-31.61)		(-31.55)		(11.86)
Peer/3W × Treated		0.02		0.02		0.02
		(0.87)		(0.87)		(0.87)
Self/3W × Treated		-0.01		-0.01		-0.01
		(-0.22)		(-0.21)		(-0.22)
Sunk-cost/3W × Treated		-0.03		-0.03		-0.03
		(-1.08)		(-1.08)		(-1.08)
Observations	443,803	286,810	443,803	286,810	443,803	286,810
Number of subjects	10,321	9,890	10,321	9,890	10,321	9,890
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: Logit model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B4: Effects on *Number of connections*: **Post-treatment** vs. **Pre-treatment**

	Number of connections					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-0.35***		-0.36***		-0.45***	
	(-12.16)		(-12.35)		(-9.86)	
Peer/1W × Treated	-0.04		-0.04		-0.04	
	(-0.90)		(-0.88)		(-0.90)	
Self/1W × Treated	0.06		0.06		0.06	
	(1.11)		(1.11)		(1.11)	
Sunk-cost/1W × Treated	0.09		0.09		0.09	
	(1.73)		(1.70)		(1.74)	
Treated × Post		-0.26***		-0.27***		-0.29**
		(-7.54)		(-7.76)		(-2.92)
Peer/3W × Treated		-0.08		-0.08		-0.08
		(-1.26)		(-1.27)		(-1.26)
Self/3W × Treated		-0.09		-0.09		-0.09
		(-1.56)		(-1.55)		(-1.56)
Sunk-cost/3W × Treated		-0.04		-0.04		-0.04
		(-0.72)		(-0.68)		(-0.72)
Constant	4.04***	4.14***	4.11***	4.28***	-72.71**	-13.11
	(278.96)	(277.06)	(160.42)	(142.19)	(-2.58)	(-0.22)
Observations	207,195	142,656	207,195	142,656	207,195	142,656
Number of subjects	10,736	10,398	10,736	10,398	10,736	10,398
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B5: Effects on *Steps walked*: **Post-treatment** vs. **Treatment**

	Steps walked					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-344.28***		-351.99***		-767.41***	
	(-9.07)		(-9.42)		(-17.24)	
Peer/1W × Treated	-50.02		-58.18		-47.54	
	(-0.78)		(-0.92)		(-0.74)	
Self/1W × Treated	-63.13		-62.21		-62.98	
	(-0.96)		(-0.96)		(-0.96)	
Sunk-cost/1W × Treated	42.54		43.45		42.72	
	(0.64)		(0.66)		(0.64)	
Treated × Post		452.86***		428.36***		1368.26***
		(14.19)		(13.61)		(29.72)
Peer/3W × Treated		-39.08		-36.59		-38.15
		(-0.70)		(-0.67)		(-0.69)
Self/3W × Treated		-92.66		-85.94		-95.48
		(-1.65)		(-1.56)		(-1.71)
Sunk-cost/3W × Treated		-40.41		-26.00		-41.07
		(-0.73)		(-0.47)		(-0.74)
Constant	6143.92***	5765.60***	6142.66***	5737.84***	-5.35e+05***	1.17e+06***
	(290.38)	(447.08)	(196.33)	(211.54)	(-17.92)	(27.66)
Observations	166,884	164,575	166,884	164,575	166,884	164,575
Number of subjects	10,107	10,081	10,107	10,081	10,107	10,081
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B6: Effects on *Steps walked*: **Post-treatment** vs. **Pre-treatment**

	Steps walked					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-517.93***		-534.41***		-1068.51***	
	(-17.83)		(-18.68)		(-23.49)	
Peer/1W × Treated	-41.25		-40.93		-39.85	
	(-0.83)		(-0.84)		(-0.80)	
Self/1W × Treated	-11.60		-14.87		-10.75	
	(-0.23)		(-0.30)		(-0.21)	
Sunk-cost/1W × Treated	-31.85		-40.90		-30.74	
	(-0.62)		(-0.81)		(-0.60)	
Treated × Post		-181.71***		-205.19***		558.98***
		(-5.42)		(-6.20)		(5.85)
Peer/3W × Treated		-61.69		-62.68		-61.83
		(-1.05)		(-1.08)		(-1.05)
Self/3W × Treated		11.93		13.58		11.07
		(0.20)		(0.23)		(0.19)
Sunk-cost/3W × Treated		-10.83		-3.31		-11.73
		(-0.18)		(-0.06)		(-0.20)
Constant	6311.31***	6355.87***	6385.69***	6515.58***	-4.36e+05***	486193.61***
	(434.92)	(437.44)	(252.85)	(224.93)	(-15.49)	(8.39)
Observations	206,760	142,353	206,760	142,353	206,760	142,353
Number of subjects	10,737	10,399	10,737	10,399	10,737	10,399
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B7: Effects on *Weekly Steps*: **Post-treatment** vs. **Treatment**

	Weekly steps					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-404.77***		-404.72***		-1084.73***	
	(-8.30)		(-8.29)		(-18.43)	
Peer/1W × Treated	26.34		26.18		28.36	
	(0.31)		(0.30)		(0.33)	
Self/1W × Treated	6.29		6.12		9.41	
	(0.08)		(0.08)		(0.12)	
Sunk-cost/1W × Treated	88.16		90.03		85.72	
	(1.02)		(1.05)		(0.99)	
Treated × Post		481.80***		478.47***		1,056.13***
		(10.36)		(10.30)		(15.92)
Peer/3W × Treated		-23.07		-23.12		-22.81
		(-0.29)		(-0.29)		(-0.29)
Self/3W × Treated		-131.86		-128.41		-131.57
		(-1.45)		(-1.41)		(-1.45)
Sunk-cost/3W × Treated		-60.24		-65.10		-63.34
		(-0.64)		(-0.69)		(-0.67)
Constant	5982.48***	5561.71***	5981.76***	5569.09***	-8.77e+05***	751492.03***
	(243.54)	(435.49)	(139.08)	(148.92)	(-19.60)	(12.86)
Observations	28,190	27,770	28,190	27,770	28,190	27,770
Number of subjects	5,638	5,554	5,638	5,554	5,638	5,554
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B8: Effects on *Weekly Steps*: **Post-treatment** vs. **Pre-treatment**

	Weekly steps					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-567.76***		-570.87***		-1551.50***	
	(-13.50)		(-13.58)		(-23.20)	
Peer/1W × Treated	32.67		35.28		30.52	
	(0.40)		(0.43)		(0.37)	
Self/1W × Treated	9.04		10.65		7.09	
	(0.13)		(0.15)		(0.10)	
Sunk-cost/1W × Treated	-31.93		-29.78		-33.93	
	(-0.41)		(-0.39)		(-0.44)	
Treated × Post		-197.73***		-201.82***		-663.62***
		(-4.18)		(-4.27)		(-5.38)
Peer/3W × Treated		-49.25		-48.10		-49.11
		(-0.61)		(-0.60)		(-0.61)
Self/3W × Treated		-47.10		-47.27		-45.70
		(-0.51)		(-0.51)		(-0.49)
Sunk-cost/3W × Treated		-64.76		-65.12		-63.66
		(-0.68)		(-0.69)		(-0.67)
Constant	6166.54***	6230.98***	6143.69***	6215.04***	-7.99e+05***	-2.98e+05***
	(332.11)	(384.81)	(167.65)	(159.53)	(-19.32)	(-4.02)
Observations	33,828	22,216	33,828	22,216	33,828	22,216
Number of subjects	5,638	5,554	5,638	5,554	5,638	5,554
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B9: Effects on *Daily Steps*: **Post-Treatment** vs. **Treatment**

	Daily steps					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-65.24		-89.53		-598.25***	
	(-0.42)		(-0.59)		(-3.32)	
Peer/1W × Treated	-229.29		-229.29		-229.29	
	(-0.83)		(-0.85)		(-0.83)	
Self/1W × Treated	-98.31		-98.31		-98.31	
	(-0.36)		(-0.37)		(-0.36)	
Sunk-Cost/1W × Treated	-231.49		-231.49		-231.49	
	(-0.84)		(-0.85)		(-0.84)	
Treated × Post		813.48***		780.39***		1815.47***
		(6.56)		(6.38)		(10.11)
Peer/3W × Treated		-881.69***		-881.69***		-881.69***
		(-3.95)		(-4.00)		(-3.96)
Self/3W × Treated		-267.58		-267.58		-267.58
		(-1.19)		(-1.21)		(-1.19)
Sunk-Cost/3W × Treated		-530.46*		-530.46**		-530.46*
		(-2.55)		(-2.58)		(-2.55)
Constant	8333.75***	8091.77***	8526.25***	8192.44***	-6.73e+05***	1.29e+06***
	(92.76)	(157.73)	(65.17)	(75.63)	(-5.59)	(7.74)
Observations	12,384	12,780	12,384	12,780	12,384	12,780
Number of subjects	344	355	344	355	344	355
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B10: Effects on *Daily Steps*: **Post-Treatment** vs. **Pre-Treatment**

	Daily steps					
	(1)	(2)	(3)	(4)	(5)	(6)
Treated × Post	-389.72**		-415.99***		-998.74***	
	(-3.24)		(-3.52)		(-5.32)	
Peer/1W × Treated	-430.92*		-430.92*		-430.92*	
	(-2.00)		(-2.04)		(-2.00)	
Self/1W × Treated	-193.60		-193.60		-193.60	
	(-0.90)		(-0.92)		(-0.90)	
Sunk-cost/1W × Treated	-184.56		-184.56		-184.56	
	(-0.85)		(-0.87)		(-0.85)	
Treated × Post		97.49		71.37		795.10*
		(0.71)		(0.53)		(2.01)
Peer/3W × Treated		-1054.22***		-1054.22***		-1054.22***
		(-4.28)		(-4.33)		(-4.28)
Self/3W × Treated		-111.37		-111.37		-111.37
		(-0.45)		(-0.46)		(-0.45)
Sunk-cost/3W × Treated		-597.20**		-597.20**		-597.20**
		(-2.60)		(-2.63)		(-2.60)
Constant	8706.62***	8826.22***	8827.61***	8992.22***	-4.82e+05***	460494.79
	(135.41)	(139.83)	(80.74)	(72.28)	(-4.15)	(1.91)
Observations	14,792	10,295	14,792	10,295	14,792	10,295
Number of subjects	344	355	344	355	344	355
Day of the week			Yes	Yes		
Time trend					Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5 : one-week arms. Columns 2-4-6 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*.

Table 3.B11: Effects on *Weekly Steps* by initial level of walking: **Post-treatment** vs. **Treatment**

	Weekly steps							
	Inactive		Somehow active		Active		Very active	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated × Post	-46.93		272.99**		99.61		-235.94	
	(-0.54)		(3.06)		(0.98)		(-1.60)	
Peer/1W × Treated	142.73		-172.66		85.40		50.15	
	(1.08)		(-1.16)		(0.52)		(0.22)	
Self/1W × Treated	242.34		-178.27		-284.64		229.51	
	(1.89)		(-1.28)		(-1.85)		(1.14)	
Sunk-cost/1W × Treated	226.50		13.29		-7.12		145.24	
	(1.69)		(0.08)		(-0.04)		(0.65)	
Treated × Post		84.62		292.43**		146.94		-129.36
		(1.01)		(3.25)		(1.48)		(-0.89)
Peer/3W × Treated		-65.05		-29.15		232.38		-255.57
		(-0.60)		(-0.22)		(1.59)		(-1.17)
Self/3W × Treated		-58.64		-251.22		-230.96		-23.54
		(-0.46)		(-1.74)		(-1.37)		(-0.10)
Sunk-cost/3W × Treated		-11.40		-197.40		-30.06		-17.46
		(-0.09)		(-1.53)		(-0.21)		(-0.06)
Constant	3233.12***	3287.81***	4772.66***	4745.75***	6229.73***	6146.44***	9756.06***	9929.58***
	(84.08)	(82.30)	(111.44)	(112.57)	(133.21)	(127.82)	(152.51)	(143.53)
Observations	7,030	6,625	7,170	7,175	7,030	6,970	6,960	7,000
Number of subjects	1,406	1,325	1,434	1,435	1,406	1,394	1,392	1,400

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5-7 : one-week arms. Columns 2-4-6-8 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*. All specifications include weekly dummies.

Table 3.B12: Effects on *Weekly Steps* by initial level of walking: **Post-treatment** vs. **Pre-treatment**

	Weekly steps							
	Inactive		Somehow active		Active		Very active	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated × Post	596.11***		300.78***		-190.87*		-1171.11***	
	(8.86)		(3.63)		(-2.32)		(-8.73)	
Peer/1W × Treated	-99.05		98.44		134.00		-24.09	
	(-1.04)		(0.79)		(1.12)		(-0.10)	
Self/1W × Treated	36.68		-130.12		-24.22		94.61	
	(0.38)		(-1.23)		(-0.20)		(0.56)	
Sunk-cost/1W × Treated	-19.63		120.29		-79.07		-100.04	
	(-0.20)		(1.01)		(-0.63)		(-0.51)	
Treated × Post		659.00***		350.04***		-248.86**		-1008.08***
		(8.99)		(4.03)		(-2.91)		(-6.82)
Peer/3W × Treated		-71.95		-66.20		81.57		-117.43
		(-0.62)		(-0.51)		(0.60)		(-0.55)
Self/3W × Treated		6.58		-96.31		15.74		11.68
		(0.06)		(-0.76)		(0.11)		(0.04)
Sunk-cost/3W × Treated		-85.82		-88.69		-46.52		29.87
		(-0.71)		(-0.73)		(-0.36)		(0.10)
Constant	2731.69***	2716.12***	4660.27***	4644.63***	6474.10***	6532.54***	10781.90***	10764.42***
	(89.57)	(92.60)	(118.03)	(130.32)	(148.44)	(159.07)	(154.05)	(161.85)
Observations	8,436	5,300	8,604	5,740	8,436	5,576	8,352	5,600
Number of subjects	1,406	1,325	1,434	1,435	1,406	1,394	1,392	1,400

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5-7 : one-week arms. Columns 2-4-6-8 : three-weeks arms. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*. All specifications include weekly dummies.

Chapter 4

Walking for Rewards: Gender Differences in Response to App-Based Incentives [1]

You're in the mood for a walk. And when you collect your Wards. You are the Walking Queen.

[1]This chapter is a joint work with Beatrice Braut (Department of Economics, University of Genoa, Italy) and Sarah Zaccagni (Department of Economics and Business Economics, Aarhus University, Denmark).

4.1 Introduction

Females are almost 18% more physically inactive than males (Strain *et al.*, 2024), leading to an increased risk of mortality by at least 20%.[2] It also contributes to mental health problems such as anxiety and stress, from which women generally suffer more (Matud, 2004; Asher *et al.*, 2017). Exercise has been found to lessen these issues significantly (Sharon-David & Tenenbaum, 2017; Bueno-Antequera & Munguia-Izquierdo, 2020) and even a simple (and free) physical activity like walking is effective in reducing stress level (Matzer *et al.*, 2018).[3] Tackling these issues by encouraging individuals to engage in regular activity is economically essential, as mental disorders considerably shorten life expectancy (Chen *et al.*, 2023) and scientists expect the associated economic burden to double by 2030 (Bueno-Antequera & Munguia-Izquierdo, 2020). Moreover, in France, encouraging only 5% of the currently inactive population to engage in regular physical activity could result in healthcare savings of 6.4 million euros (ISCA & CEBR, 2015). In 2022, 37% of the French population declared using a health app and, in 2023, 73% of U.S respondents reported using a health tracking app at least once a week.[4,5] Given the increasing use of activity monitoring applications and their low associated cost, health policies are gradually relying on health apps – a practice known as mHealth.[6] In their *Global Action Plan on Physical Activity*, the World Health Organization (2019) underscores their role in public health policy: "*build upon the rapidly growing practice of mHealth to harness the potential of data to help promote, support and monitor physical activity*" (p.18).

Given the context, we investigate possible gender differences regarding health app fea-

[2]See the World Health Organization's website: https://www.who.int/news-room/fact-sheets/detail/physical-activity

[3]A study from Broocks *et al.* (1998) demonstrates that endurance training produces an effect similar to that of antidepressants on anxiety and depressive disorders.

[4]For the share of health app users in selected countries (2022), see https://www.statista.com/forecasts/1181436/share-of-health-app-users-by-country

[5]For the distribution of frequency of health tracking app usage in a month in the U.S (2023), see https://www.statista.com/statistics/1424311/frequency-of-mhealth-use-in-the-us/

[6]The term refers to the use of mobile devices (such as mobile phones or smartwatches) to collect health data, share healthcare information or monitor patients' vital signs in real-time.

tures. More specifically, we test the effect of monetary prizes on engagement through the app. Our data emerge from a large study conducted in collaboration with the step-tracking app *WeWard* that motivates its users to walk through monetary prizes. Previous research suggests that men are more responsive to monetary incentives, compared to non-monetary ones, regarding school performance (Levitt *et al.*, 2016) or real-effort task (Sittenthaler & Mohnen, 2020). Besides, Alexander *et al.* (2008) demonstrates that financial incentives to promote enrollment in an online health program are more effective among men than women. Our chapter focuses on the use of a health app and we aim to answer the following research question. Are men and women equally sensitive to the monetary incentives proposed by health apps?

Many features participate to the success of health apps, such as gamification (King *et al.*, 2013), reminders (Free *et al.*, 2009; Lester *et al.*, 2010), social network (Christakis & Fowler, 2007; Centola, 2013; Tu *et al.*, 2019; Franken *et al.*, 2023), or non-monetary rewards (Feng *et al.*, 2020; Yin *et al.*, 2022; Faizah *et al.*, 2024). With nearly 41,517 health apps available on the AppleStore, apps are striving to stand out by offering new and innovative features.[7] Recently, some applications propose financial incentives since they have been found effective in promoting exercise (Charness & Gneezy, 2009; Acland & Levy, 2015; Vlaev *et al.*, 2019). Monetary incentives reduce the costs of starting a physical activity, which may be perceived by individuals as exceeding the short-term benefits (Andersen, 1999). However, the efficacy of monetary incentives remains limited in the long term when focusing on activities requiring maintenance costs, such as gym attendance (Charness & Gneezy, 2009; Carrera *et al.*, 2018). Therefore, digital apps that continually offer monetary incentives can help build or reinforce healthy habits. For instance, Rondina *et al.* (2021) demonstrates the cost-effectiveness of *Carrot Rewards*, an app encouraging Canadians to walk through monetary prizes. However, the authors construct a model for assessing the benefits and costs of this type of application without really examining the interaction of users with the app. We believe in the importance of examining how users interact with these apps in order to promote specific features based

[7]For the number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 3rd quarter 2022, see https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/

on individual characteristics.

Few studies assess gender differences in the use of health applications, despite their potentially large impact on the implementation of health policies using this tool. Previous research reports gender differences in app utilization of *Runtastic* for instance, an application combining social channels and gamification elements (Klenk *et al.*, 2017). According to Klenk *et al.* (2017), women value enjoyment and achieving goals while men enhance social comparison by sharing more often their results. This pattern could explain why men download more *Strava* (an app allowing peer comparison) compared to women (Jestico *et al.*, 2016).[8] Regarding platform rewards, papers usually focus on non-monetary ones, such as badges or likes (Feng *et al.*, 2020; Yin *et al.*, 2022; Faizah *et al.*, 2024). Yin *et al.* (2022) and Faizah *et al.* (2024) both found that women are more sensitive to external factors (e.g., social support) whereas men are more motivated by rewards. A closely related study to our research is by Yeung *et al.* (2024), in which the authors conducted a randomized controlled trial (RCT) to investigate gender differences in response to financial incentives for weight loss. Results show that monetary incentives are effective for males but not for females.

Our research investigates gender differences in the use of a step-tracking app. Regarding recorded walking behaviors, our data suggest that women walk significantly less than men. However, they open the application significantly more often compared to men. Women also appear more committed to the application in terms of steps-to-prizes conversion. Nevertheless, men and women do not significantly differ in how they respond to monetary prizes. Both tend to interact with *WeWard* conditionally to the reward received from the app.

The rest of the chapter is organized as follows. Section 4.2 describes the main *WeWard* application features. The results are presented in Section 4.3. Section 4.4 concludes the chapter.

[8]More broadly, men are more likely to download sports-related apps compared to women (see "Are Men and Women Using Mobile Apps Differently?", Wired (2013): https://www.wired.com/2013/04/men-women-app-usage/)

4.2 The app features

We Ward is a step-tracking application that was launched in 2019 following public transportation strikes in France and has since been downloaded by one in five French citizens.[9] *We Ward* is an application that motivates individuals to engage in walking activities through monetary prizes. The app works as a pedometer and retrieves users' daily steps when the user opens the app. After opening the app, the user can convert their steps into points (named *Wards*) by clicking on the "Valider mes pas" button, as presented in Figure 4.A1 (Appendix 4.A). Users can convert their steps into *Wards* (henceforth steps-to-*Wards* conversion) at any time during the day but they need to validate their steps before midnight, otherwise the points are lost for that day. More importantly, users must reach a certain threshold of steps to earn points. The thresholds are: 1,500, 3,000, 6,000, 10,000, 15,000, and 20,000 steps and the steps between two thresholds do not give extra *Wards*. After reaching a certain number of *Wards*, corresponding to 15 €, users can convert their points into cash.

4.3 Results

4.3.1 Descriptive statistics

We benefit from a large dataset of users across France (N = 14,215).[10] The period of observation lasts 7 weeks: from November, 29th 2022 to January, 17th 2023.[11] In addition, to provide a more detailed overview, we also exploit administrative data from INJEP *et al.* (2020) and

[9] https://www.decideurs-magazine.com/strategie/56508-tanguy-de-villegeorges-weward-un-francais-sur-cinq-a-telecharge-l-application.html

[10] Our initial database is composed of 20,187 individuals but since providing their gender is not mandatory on the application, 3,986 users do not fill in this information. We drop them from the analysis. Additionally, 2,212 individuals never opened the application and 598 users have a null amount of steps during the observed period.

[11] As presented in Figure 4.A2 in Appendix 4.A, our sample covers all the Metropolitan and overseas French regions and the proportion (per thousand) of users relative to the population is somehow homogeneous. The overseas territories correspond to the DROM-COM, which stands for "Départements et régions d'outre-mer, Collectivités d'outre-mer" (Mayotte, Martinique, Guyane and La Réunion).

INJEP & MEDES (2020).[12] Overall, we find that women declare to walk more than men but when looking at *WeWard*'s recorded steps, the gap is reversed. For instance, women are significantly less likely to record 10,000 steps (i.e. WHO recommendation) compared to men. Moreover, we observe gender differences in exercise motivations. According to our administrative data, men report competition and improvement in personal performance as reasons for engaging in physical activity more often than women, as shown in Table 4.B3 (Appendix 4.B.3). On the other hand, women are more sensitive to health (such as reduction of stress) and physical objectives.

Regarding app usage, women open significantly more the app than men. As shown in Table 4.B1 (Appendix 4.B.2), on average, 51.50% of our female sample open the app at least one time per day against 44.08% of males (χ^2, $p = 0.00$). Figure 4.1 illustrates the percentage of users who open the app by gender and day of the week. It demonstrates that the difference accounts for each day of the week.[13] Based on app usage frequency, women seem to be more engaged with *WeWard*. In the subsequent section, we delve into user steps-to-*Wards* conversion behaviors to investigate whether men exhibit a stronger response to the app's monetary incentives.

4.3.2 Investigation of threshold's effect

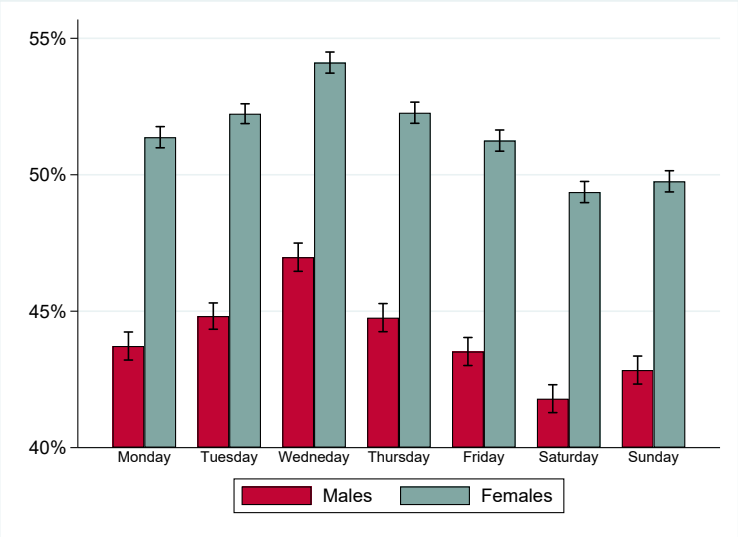
Figure 4.2 illustrates the share of users who validate their steps in relation to the step thresholds and using kernel smoothing method.[14] The red dotted bars represent the step thresholds for earning more points on the application: 1,500, 3,000, 6,000, 10,000, 15,000,

[12]The descriptive statistics of recorded and self-reported walking habits are presented in Appendix 4.B.1. Refer to Appendix 4.B.3 for details regarding the respective administrative database and the summary tables reporting the main variables of interest.

[13]The weekly pattern appears to be similar for men and women: the share of subjects who open the app increases until Wednesday before decreasing over the weekend. Indeed, both genders are less likely to open *WeWard* during the weekends, compared to the week's days (χ^2, $p = 0.00$). Moreover, males and females share similar trends over the studied period as shown in Figure 4.C1 (Appendix 4.C.1).

[14]We restrict the sample to those who walk between 1,000 and 20,500 steps to have a cleaner illustration of the jumps near to the thresholds.

Figure 4.1: Share of users who open *We Ward*, by day of the week

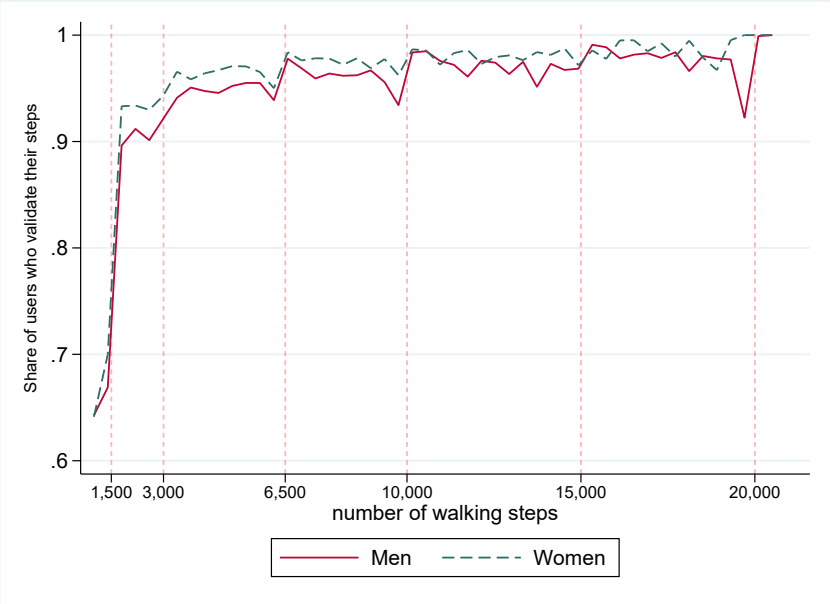


Notes: Confidence intervals at 95%.

and 20,000. For each threshold of steps, on average, females are significantly more likely to validate their steps compared to males (χ^2, $p = 0.00$). The share of users who validate their steps strongly increases for both genders when the number of steps walked approaches the first threshold. The peaks are not exactly at the threshold level since users do not usually walk precisely the amount of the threshold. Men's curve appears more fluctuating, with more pronounced peaks compared to women's. Close to the last threshold, men seem to be strongly influenced by reaching the threshold, as the share of male users who validate their steps drops before highly increases. On the other hand, female's pattern of validation is quite stable across the number of steps.

We further analyze the steps-to-*Wards* conversion behaviors using a Multi-Cutoff Regression Discontinuity Design (RDD). The RDD method allows investigating the effect of a specific cutoff on the probability of validating steps (takes the value 1 if the subject walks and validates at least some steps and 0 if the user walks but does not validate their steps) for users near the cutoff (i.e. below or above). Regarding our data, a sharp design appears pertinent since it corresponds to the following case: "*every subject is assigned a score and a treatment is given to all units who score is above the cutoff and withheld from all units whose*

Figure 4.2: Probability of validating steps in relation to the thresholds, by gender



score is below it" (p.1229, Cattaneo *et al.*, 2016). In our case, the score corresponds to the number of walking steps, the cutoffs are the thresholds, and the treatments are the additional *Wards* earned to walk above the threshold. The chosen proximity to the cutoff is arbitrary 500 steps, meaning that we consider users above/below the thresholds by 500 steps.[15]

In order to properly compare users near the thresholds (within 500 steps), we draw an RDD plot by merging all the thresholds into a single cutoff, as illustrated in Figure 4.3.[16] Figure 4.3a presents the male pattern of validation and Figure 4.3b focuses on the female sample. Figure 4.3b demonstrates that women tend to validate more after reaching the threshold but they tend to maintain a relatively stable curve, even before reaching the cutoff. In contrast, men's probability of validation increases upon reaching the threshold but slightly decreases when they fall short of it. Figures 4.2 and 4.3 suggest that men experience a larger increase in their probability of validating steps after reaching the threshold, compared to women. We perform Multi-Cutoff RDD regressions to further investigate whether men are more sensitive to a bunching effect. A bunching effect would correspond to men's tendency

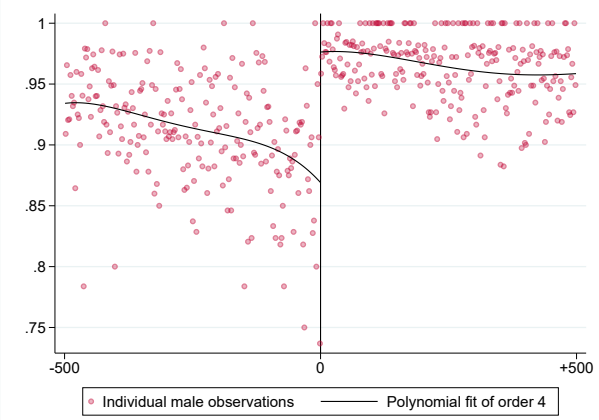
[15]We conduct robustness checks in Tables 4.C1 and 4.C2 in Appendix 4.C.2 with a cutoff respectively to 200 and 700 steps near the threshold. The results are robust.

[16]Additionally, we present RDD plots for each threshold in Figures 4.C2 to 4.C7 (Appendix 4.C.1).

to adjust their step conversion behavior to cluster around thresholds, giving evidence that they are more sensitive to the prize thresholds compared to women.

Figure 4.3: RDD plots for all thresholds

(a) Males



(b) Females

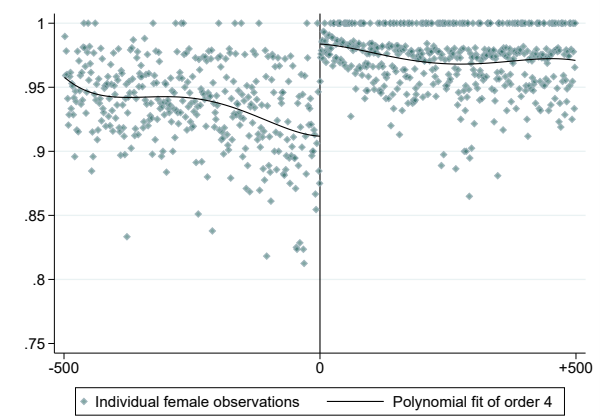


Table 4.1 presents two specifications using a Multi-Cutoff RDD estimator on the probability of validating steps. Column 1 focuses on our male sample and column 2 assesses the female sample. The coefficients of the thresholds report the effect of the discontinuity, i.e. the estimated effect of threshold crossing. The weighted and pooled coefficients display the results of the combined estimation across all thresholds.[17] In our case, the weighted estimator gives greater weight to the threshold (here 6,500) with the largest number of observations.

Significant positive coefficients in both columns suggest that men and women are more likely to validate steps after crossing a threshold. However, the effect diminishes and becomes insignificant at the higher thresholds. Additionally, the coefficients for men are higher than those for women. When crossing the threshold, men's probability of validating steps increases by 0.15 units on average against 0.12 for women. This result suggests that men are more influenced by the threshold in their validation behavior. We examine the confidence intervals for the estimated coefficients by gender in Figure 4.C8 (Appendix 4.C.1). The difference in the coefficients of men and women does not appear significant. Therefore, we cannot state

[17]Tables 4.C3 and 4.C4 (Appendix 4.C.2) present the results of sharp RD estimator for the male and female sample respectively. The results are robust.

gender differences regarding responses to prize thresholds. Both male and female interact more with the application *WeWard* when being rewarded.

Table 4.1: The effects of thresholds on the probability of validating steps

	Probability of validating steps	
	Males	Females
	(1)	(2)
Threshold = 1,500	0.25***	0.19***
	(5.93)	(9.71)
Threshold = 3,000	0.11***	0.10***
	(3.81)	(7.13)
Threshold = 6,500	0.08***	0.05***
	(3.52)	(4.04)
Threshold = 10,000	0.12**	0.09***
	(3.10)	(3.44)
Threshold = 15,000	0.04	-0.04*
	(0.82)	(-2.14)
Threshold = 20,000	0.29	-0.03
	(1.66)	(-0.99)
Weighted	0.13***	0.11***
	(8.40)	(12.95)
Pooled	0.15***	0.12***
	(8.66)	(13.87)

Notes: Cutoff-specific RD estimation with robust bias-corrected inference. t statistics in parentheses. The dependent variable in columns 1-2 is the probability of validating the steps. Column 1: male sample and column 2: female sample. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

4.4 Conclusions

We provide a gender-focused analysis of app usage, aiming to shed light on efficient mechanisms for encouraging physical activity among men and women through health applications.

Our analysis of a step-tracking app reports no significant gender difference regarding the effect of monetary incentives on the use of the app. Crossing a step threshold (that provides additional rewards) affects both males' and females' steps-to-*Wards* conversion behavior. The effect is significant only for thresholds under 15,000 steps. This result suggests that monetary incentives are effective in encouraging users to interact with health apps. Consequently, an increase in user interest can lead to better monitoring of walking activity and then greater awareness of the level of practice.

Given the differences between self-reported and recorded measures, our study also underscores the importance of tracking actual behavior when collecting data or evaluating interventions. Reports usually measure physical activity through self-report but this tool tends to underestimate women's activity (Ainsworth, 2000; Collins *et al.*, 2007).[18] Data recorded via a pedometer system on a mobile phone circumvents self-reporting issues such as social pressure or demand-effect bias. Moreover, accessing non-declarative data helps mitigate declarative bias and underestimation of certain activities.

However, the use of a pedometer does not eliminate all the problems since motion sensors may underpredict the energy cost of certain activities (Bassett, 2000). Users might forget to carry their phone, leading to underreporting of their steps. The analysis of recorded data could also face a problem of "voluntary" attrition due to app usage strategies.[19] Health applications are becoming increasingly popular and have the potential to generate significant long-term health benefits without imposing substantial short-term costs. However, for digital health interventions, it is crucial to measure behavior beyond mere app usage. Studies should consider that users might adopt app usage strategies that could bias the recording of their physical activity.

[18]Daily tasks, such as domestic caring duties, constitute physical activity (Brown *et al.*, 2001) but they could be not considered as proper physical tasks (Fjeldsoe *et al.*, 2009).

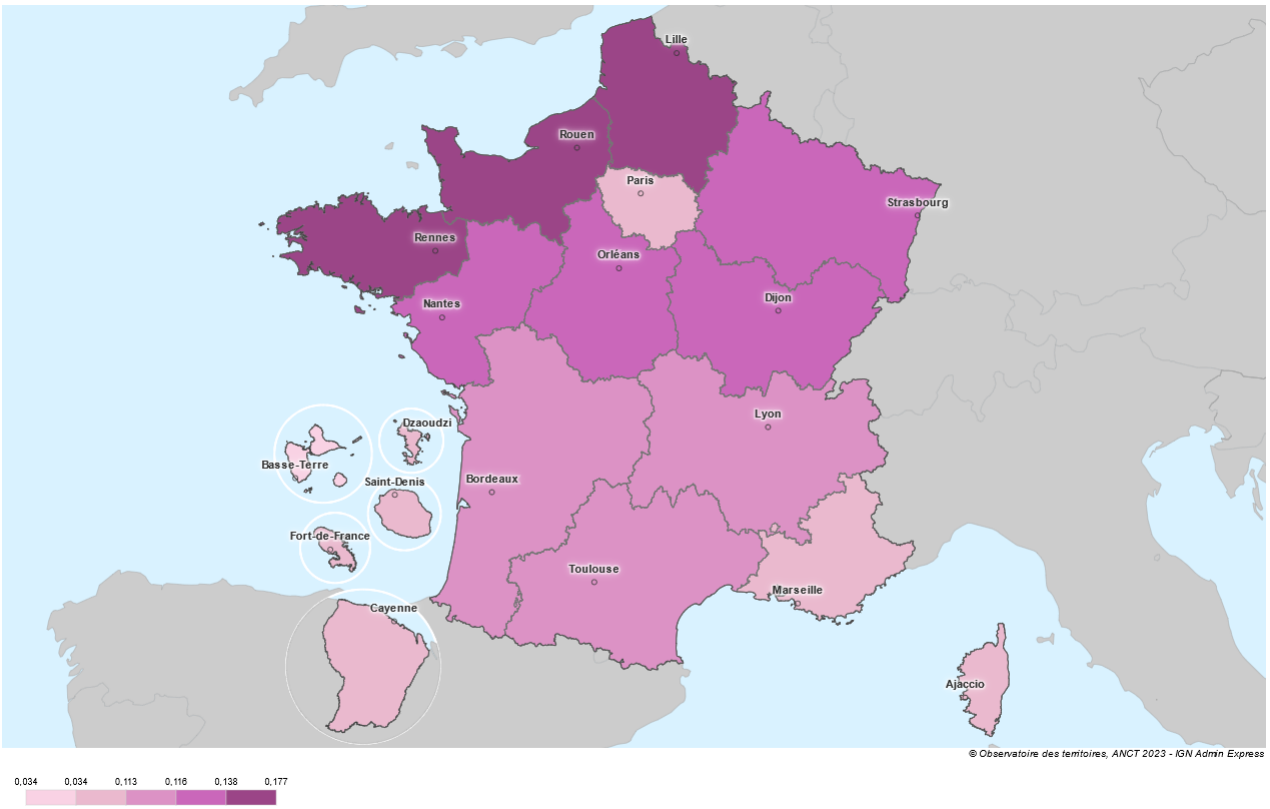
[19]In our study, given that women open more *WeWard*, their steps are more often retrieved by the app, which could lower their average step count. Since the steps are retrieved only when the user opens the app, we cannot confirm this pattern. Although it constitutes an interesting outcome, the attrition in recorded steps is one of the limits of our study.

4.A Background

Figure 4.A1: Screenshot of the app



Figure 4.A2: Fraction of users in the population, by French region



Notes: Figure 4.A2 illustrates the fraction of users (in ‰) in the population by region. Overall, the user rate is homogeneous. The lowest user rate is in DOM-TOM (i.e. 0.03‰) and the largest is in Hauts-de-France (i.e. 0.17‰).

4.B Walking behaviors

4.B.1 An overview of gender differences in walking habits

Table 4.B1 (Appendix 4.B.2) presents the summary statistics of our *WeWard* data. Among the whole sample, women record on average 5,558.31 steps per day against 7,126.72 for men (Mann-Whitney, henceforth MWU, $p = 0.00$). As shown in Figure 4.B1 (Appendix 4.B.2), women are more likely to have smaller values than men (Kolmogorov-Smirnov, $p = 0.00$). Based on the recorded steps, women are 9.51% more likely to be sedentary (i.e., less than 5,000 steps), 10.49% more to walk less than 8,000 steps (threshold below which the risk of disease increases, see Inoue *et al.*, 2023) and 9.01% more likely to not achieve the WHO recommendation of 10,000 steps per day compared to men. The gap is observed both on a weekly and daily basis. Indeed, even after restricting our database to those who open the app every day, we still observe gender differences in favor of men.[20] Females are more likely to record their steps every day (χ^2, $p = 0.00$) but they seem to walk significantly less than men (MWU, $p = 0.00$). As shown in Figure 4.B2, this difference is significant for each day of the week. In recent years, the average time to complete a marathon for women has risen, revealing an increase in the number of female participants.[21] Similarly, the average amount of steps reported by women on the app could be explained by the fact that women record lower amounts than men.

As a matter of fact, our self-reported data does not reveal gender differences disadvantaging women. As shown in Tables 4.B2 and 4.B3, women declare significantly more to walk as a sportive activity (Athletic and Nordic in Table 4.B2) or as a regular activity (Table 4.B3). As opposed to Goel *et al.* (2023), these data do not show significant differences when subjects

[20] When considering only those who open the application at least one day per week, 1,416 individuals do not fill in their gender, leading to a sample of 7,525 individuals. The sample considered for the weekly habits includes 31.60% of male observations and 68.40% of female observations. On the other hand, our restricted sample of strongly committed users is composed of 495 individuals, with 67.27% of women and 32.73% of men.

[21] "En moyenne, en combien de temps finit-on un marathon ?", Brooks (2023): *https://www.brooksrunning.com/fr_fr/blog/advice-tips/average-marathon-time.html*

declare to use walking as a transportation mean (Table 4.B2). The inconsistency between self-reported and recorded measures may be explained by a declarative bias, or by different strategies adopted by men and women in their use of the application that we explore across the chapter.

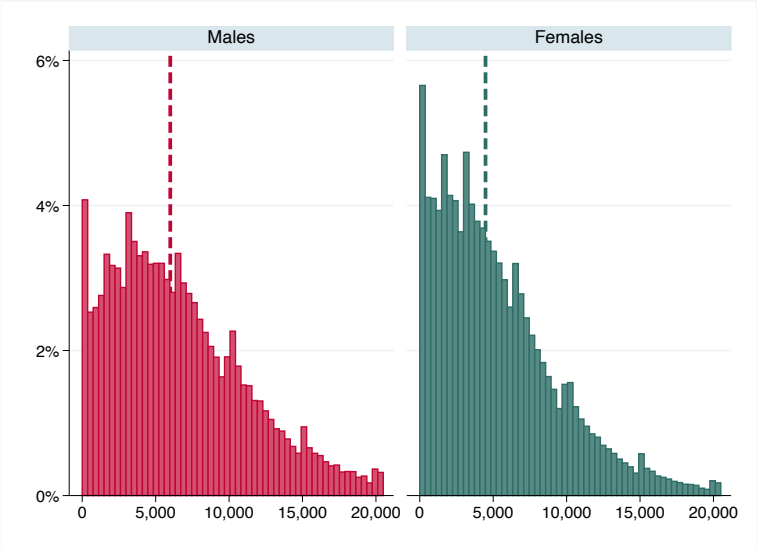
4.B.2 Recorded measures

Table 4.B1: Summary statistics *WeWard* dataset

	Females	Males	p	Total
Walking habits				
Average number of steps	5,558.31	7,126.72	0.00	6,065.10
	(4729.20)	(5618.77)		(5081.01)
% of subjects who walk $<$ 5,000 steps	0.26	0.17	0.00	0.23
	(0.44)	(0.38)		(0.42)
% of subjects who walk $<$ 8,000 steps	0.37	0.26	0.00	0.33
	(0.48)	(0.44)		(0.47)
% of subjects who walk $>$ 10,000 steps	0.22	0.31	0.00	0.25
	(0.41)	(0.46)		(0.43)
App engagement				
Probability of opening the app	0.51	0.44	0.00	0.49
	(0.50)	(0.50)		(0.50)
Probability of validating steps	0.95	0.96	0.00	0.95
	(0.31)	(0.29)		(0.31)
N	9,106	5,109		14,215
	(64.06%)	(35.94%)		

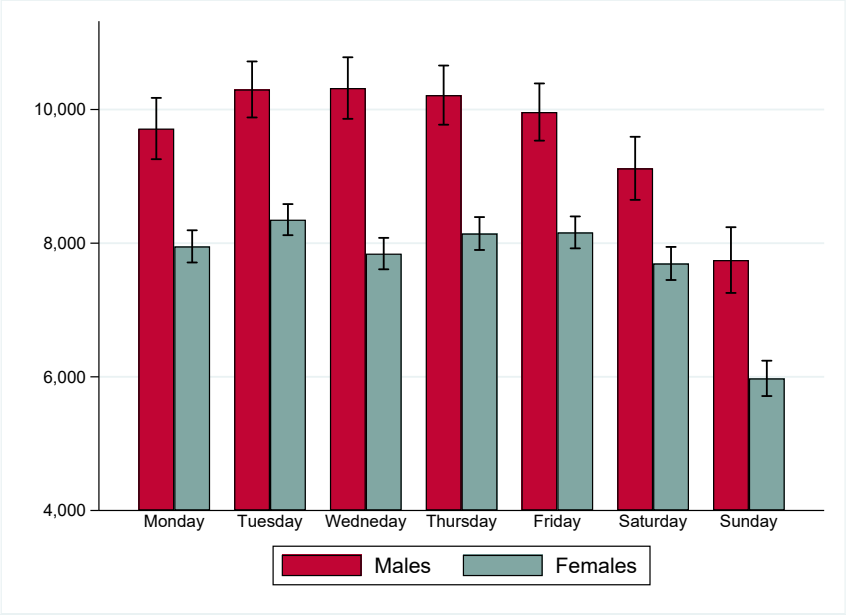
Notes: Standards deviations in parentheses. The p-value is associated with χ^2 tests for the binary variables and MWU for the categorial variables.

Figure 4.B1: Gender gap in the distribution of the number of steps



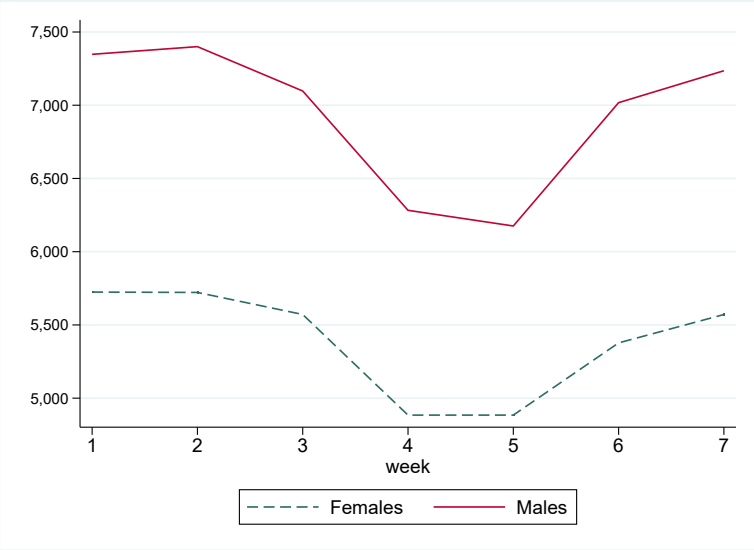
Notes: We do not consider the number of steps superior to 20,500, leading to a sample equals to 14,199. The dot bars represent the median for males (5,995) and for females (4,478).

Figure 4.B2: Gender gap in the number of recorded steps by day of the week



Notes: Confidence intervals at 95%.

Figure 4.B3: Evolution of the weekly steps across the observation weeks, by gender



4.B.3 Self-reported measures

Bâromètre de la pratique sportive - Édition 2020:

The survey was conducted by the Ministry of sports, Centre de Recherche pour l'Étude et l'Observation des Conditions de vie (CREDOC), and Institut National de la Jeunesse et de l'Éducation Populaire (INJEP). Data cover 4,079 individuals aged between 15 and 92, with 52.95% of females and 47.05% of males.[22,23]

Table 4.B2: Baromètre de la pratique sportive - Édition 2020

	Females	Males	p
Footing, jogging			
% of runners	15.88	20.58	0.00
Mean of number of times practised the last year	78.28	90.89	0.00
Average time per workout (min)	49.02	54.51	0.02
Intensity of the workout	1.63	1.67	0.38
Self-report level	2.40	2.78	0.00
Walking			
As a transportation mean (%)	23.29	25.17	0.16
Athletic and Nordic walker (%)	4.95	2.87	0.00
Other physical related activity (%)			
Users of an instrument to measure activity (watch or app)	26.11	31.37	0.00
Spectators of a sports event (e.g. match)	40.46	67.34	0.00
Users of a community application (e.g. Strava)	10.82	10.14	0.75

Notes: The intensity of the workout ranges from 1 (= moderate) to 3 (= intense). The self-reported level ranges from 1 (= beginner) to 5 (= expert). For the share of users of a community application, we only consider those who report practising jogging, athletic walking or Nordic walking. The p-value is associated with χ^2 tests for the binary variables and MWU for the categorial variables.

[22]All the measures used in Table 4.B2 refer to answers from questions where the investigator always specifies "apart from the lockdown period". Other questions, unused in this study, state "during the lockdown period".

[23]Data are available on request at https://data.progedo.fr/studies/doi/10.13144/lil-1497.

Enquête nationale sur les pratiques sportives - Édition 2020:

Data were collected by the Ministry of sports, Ministry of national education and youth, and Institut National de la Jeunesse et de l'Éducation Populaire (INJEP). Database contains 12,034 individuals with 52.57% of females and 47.43% of males.[24,25]

Table 4.B3: Enquête nationale sur les pratiques sportives - Édition 2020

	Females	Males	p
Footing, jogging			
% of runners	19.95	28.02	0.00
% road racers (10 km, half-marathon, marathon)	0.10	0.21	0.16
Walking			
% of walkers	83.25	81.02	0.02
Other physical related activity (%)			
Users of sport application	21.47	23.90	0.00
Motivation for physical activity			
Competition	-	+	0.00
Feel good about yourself, reduce stress	+	-	0.00
Improve physical appearance	+	-	0.00
Improve personal performance	-	+	0.00
Physical maintenance	+	-	0.00
Pleasure	-	+	0.00
Barriers for physical activity			
Distate for sport	+	-	0.00
Do not know anyone to practice with	+	-	0.00
Family constraints	+	-	0.00
Fear of others' gaze	+	-	0.00
Financial constraints	+	-	0.00

Notes: The p-value is associated with χ^2 tests for the binary variables and MWU for the categorial variables. The "+" corresponds to the over-representation of one gender in the answer and the "-" represents the under representation. We do not consider all the motivations or barriers asked in the survey and select those that seem most relevant.

[24] Again, all the questions used state explicitly "apart from the lockdown period".

[25] Data are available on request at https://data.progedo.fr/studies/doi/10.13144/lil-1620.

4.C App behavior

4.C.1 Additionnal figures

Figure 4.C1: Time trend of share of subjects who open *WeWard*, by gender

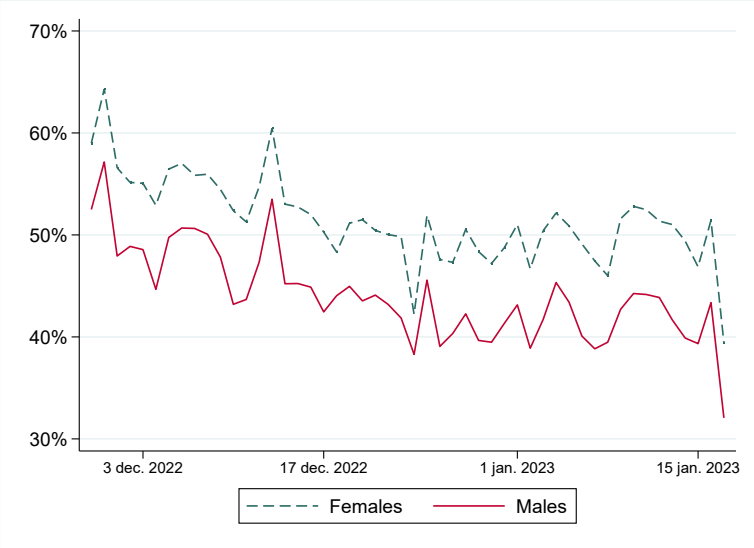
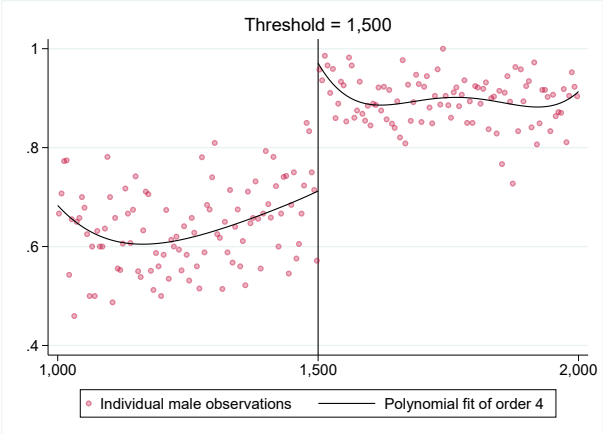


Figure 4.C2: RDD plot for the **first threshold**

(a) Males



(b) Females

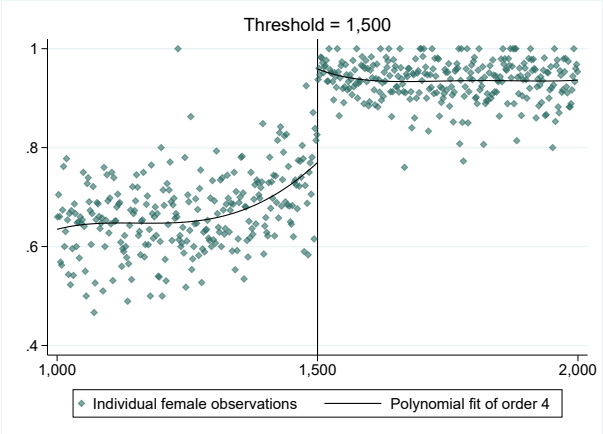
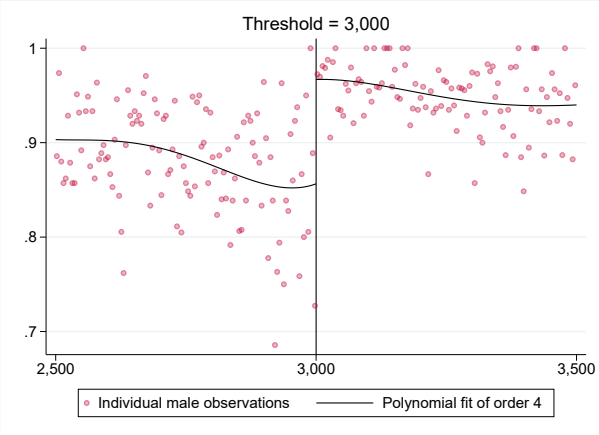


Figure 4.C3: RDD plot for the **second threshold**

(a) Males



(b) Females

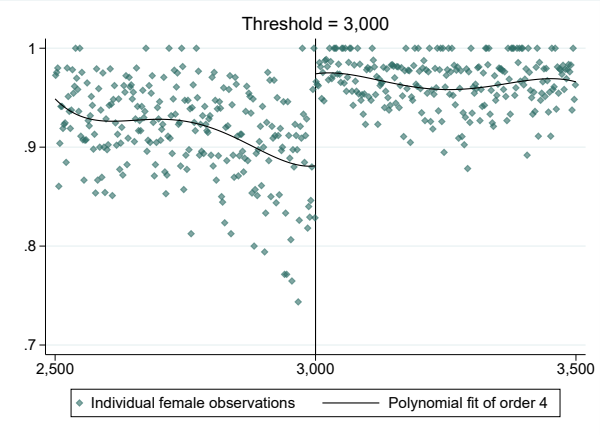
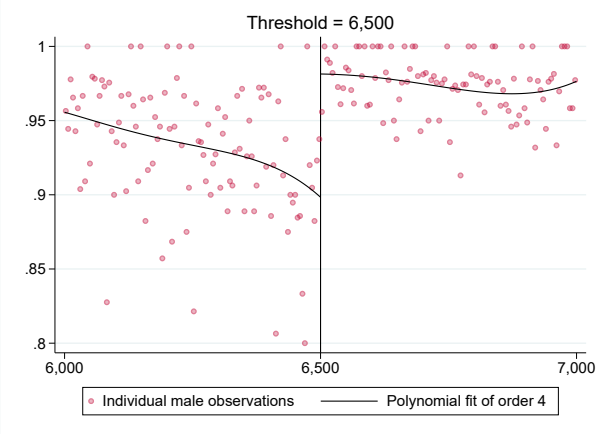


Figure 4.C4: RDD plot for the **third threshold**

(a) Males



(b) Females

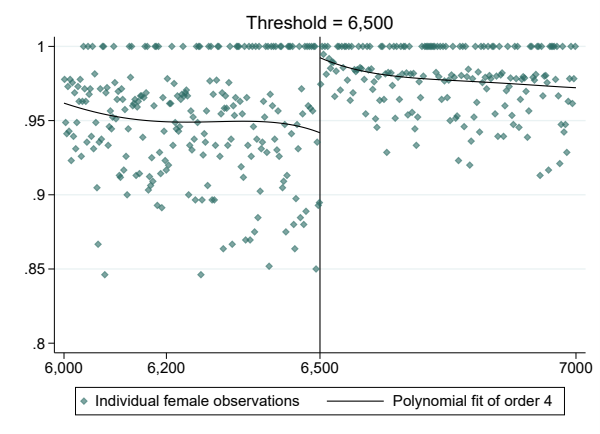
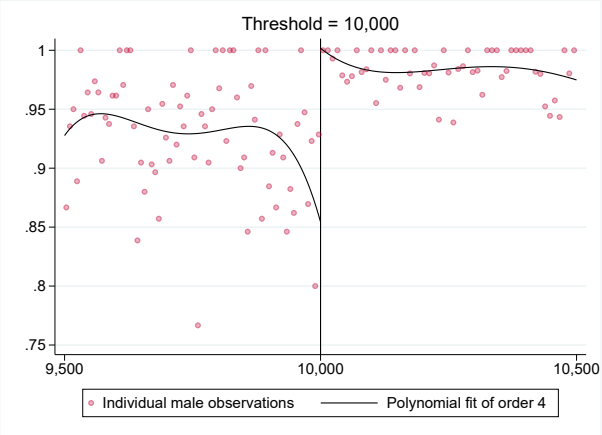


Figure 4.C5: RDD plot for the **fourth threshold**

(a) Males



(b) Females

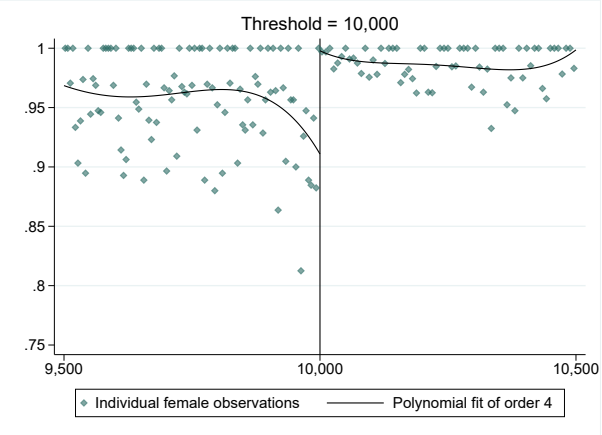
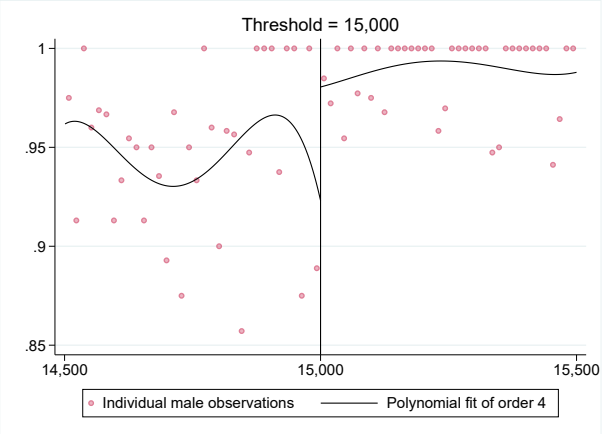


Figure 4.C6: RDD plot for the **fifth threshold**

(a) Males



(b) Females

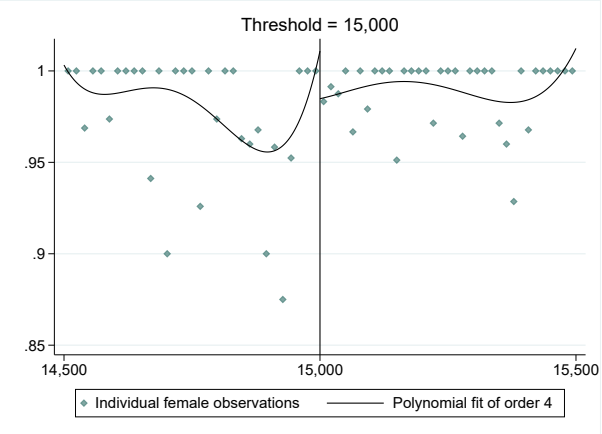
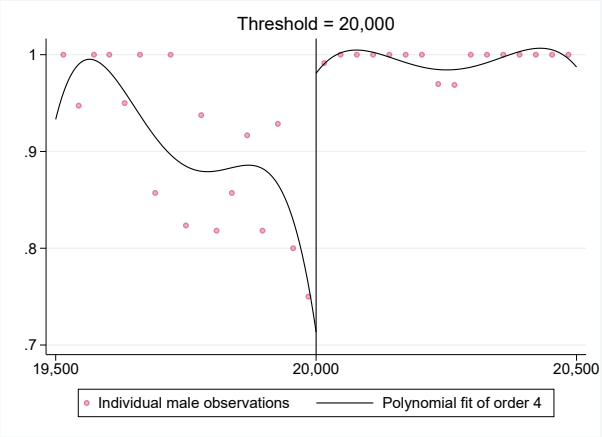


Figure 4.C7: RDD plot for the **sixth threshold**

(a) Males



(b) Females

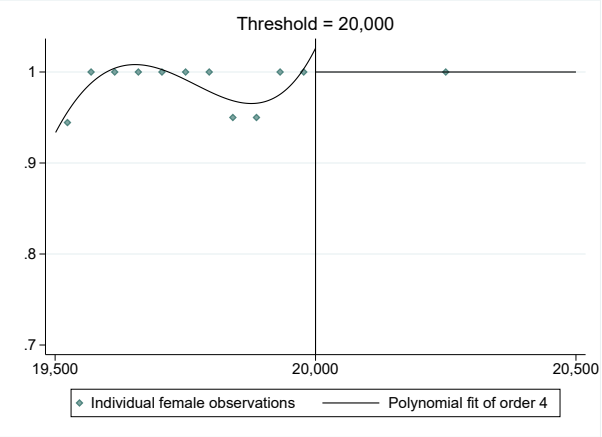
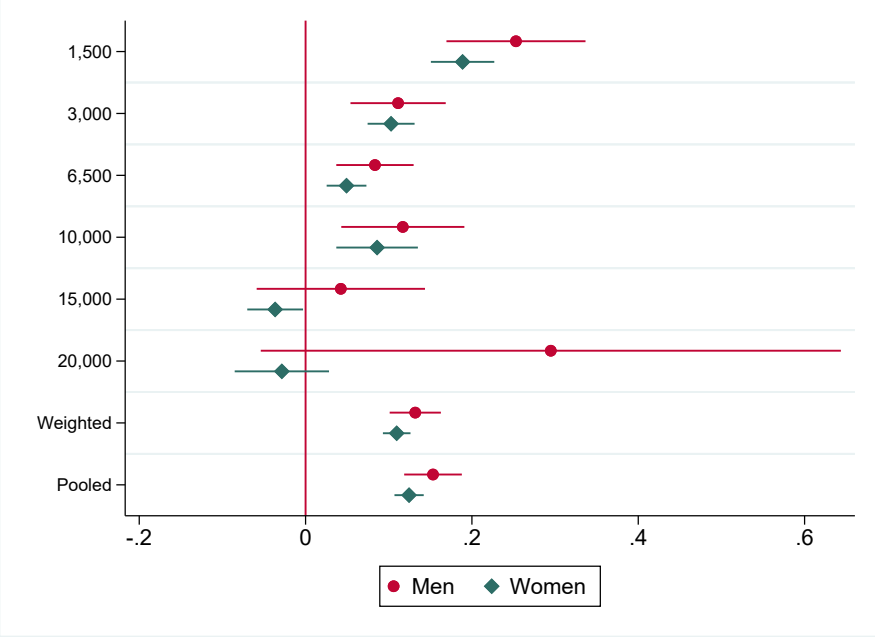


Figure 4.C8: Confidence intervals for the estimated coefficients of Table 4.1, by gender



Notes: Confidence intervals at 95%.

4.C.2 Additionnal tables

Table 4.C1: The effects of thresholds on the probability of validating steps with a cutoff to **200 steps**

	Probability of validating steps	
	Males	Females
	(1)	(2)
Threshold = 1,500	0.23***	0.18***
	(4.03)	(4.51)
Threshold = 3,000	0.13*	0.07*
	(2.15)	(2.55)
Threshold = 6,500	0.04	0.05*
	(0.88)	(1.96)
Threshold = 10,000	0.13*	0.05
	(1.98)	(1.19)
Threshold = 15,000	0.01	-0.04
	(0.09)	(-1.58)
Threshold = 20,000	0.24	
	(1.00)	
Weighted	0.13***	0.09***
	(4.72)	(5.36)
Pooled	0.16***	0.11***
	(4.93)	(5.97)

Notes: Cutoff-specific RD estimation with robust bias-corrected inference. t statistics in parentheses. The dependent variable in columns 1-2 is the probability of validating the steps. Column 1: male sample and column 2: female sample. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively. Note that, for the last threshold, the coefficient is omitted for females due to the lack of observations.

Table 4.C2: The effects of thresholds on the probability of validating steps with a cutoff to **700 steps**

	Probability of validating steps	
	Males	Females
	(1)	(2)
Threshold = 1,500	0.23***	0.19***
	(7.37)	(9.32)
Threshold = 3,000	0.11***	0.11***
	(4.63)	(8.09)
Threshold = 6,500	0.08***	0.05***
	(4.06)	(4.27)
Threshold = 10,000	0.12***	0.09***
	(3.55)	(3.45)
Threshold = 15,000	0.04	-0.04*
	(0.79)	(-2.54)
Threshold = 20,000	0.30	-0.02
	(1.66)	(-0.78)
Weighted	0.14***	0.11***
	(10.21)	(13.30)
Pooled	0.15***	0.12***
	(11.08)	(13.95)

Notes: Cutoff-specific RD estimation with robust bias-corrected inference. t statistics in parentheses. The dependent variable in columns 1-2 is the probability of validating the steps. Column 1: male sample and column 2: female sample. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

Table 4.C3: The effects of being above the threshold for males

	Probability of validating steps					
	Thresholds					
	1,500	3,000	6,500	10,000	15,000	20,000
	(1)	(2)	(3)	(4)	(5)	(6)
RD Estimator	0.25***	0.11***	0.08***	0.11***	0.04	0.26
	(6.65)	(3.85)	(4.14)	(3.42)	(0.85)	(1.73)
Observations	8,166	9,147	8,274	5,360	2,009	808

Notes: Sharp RD estimator using local polynomial regression. t statistics in parentheses and standard errors are robust and clustered at *customerID*. The dependent variable in columns 1-6 is the probability of validating the steps. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

Table 4.C4: The effects of being above the threshold for females

	Probability of validating steps					
	Thresholds					
	1,500	3,000	6,500	10,000	15,000	20,000
	(1)	(2)	(3)	(4)	(5)	(6)
RD Estimator	0.19***	0.10***	0.05***	0.08***	-0.02	-0.02
	(11.15)	(6.87)	(4.66)	(3.77)	(-1.89)	(-1.01)
Observations	24,768	24,003	16,623	8,330	2,445	929

Notes: Sharp RD estimator using local polynomial regression. t statistics in parentheses and standard errors are robust and clustered at *customerID*. The dependent variable in columns 1-6 is the probability of validating the steps. Independent variables described in text. *, **, *** indicate significance at the 5%, 1%, 0.1% level, respectively.

Conclusion générale

Cette thèse s'inscrit dans le champ de l'économie comportementale et vise à étudier l'expression comportementale qui se produit en réaction à une comparaison sociale (Wood, 1996). Notre étude rend compte des effets de la comparaison sur les comportements et perceptions des individus dans plusieurs domaines, ambitionnant ainsi de répondre à la question suivante : dans quelle mesure nos comportements face à la compétition, aux inégalités ou encore à la santé peuvent-ils être influencés par nos pairs ? Pour appréhender cette question, notre thèse mobilise les outils de l'économie expérimentale.

L'intérêt premier d'une telle méthodologie est qu'elle permet d'étudier des comportements individuels associés à des conséquences économiques. Notre premier chapitre porte sur l'esprit de compétition en fonction du genre et se place dans un contexte d'inégalités salariales entre les femmes et les hommes. Il existe deux formes de ségrégation sur le marché du travail : horizontale et verticale. La ségrégation horinzontale (ou professionnelle) rend compte du fait que la plupart des métiers féminisés sont moins rémunérateurs. En 2011, 52% des femmes auraient dû changer de profession pour mettre fin à la ségrégation horizontale en France (Argouarc'h & Calavrezo, 2013). Cette dernière demeure aujourd'hui, étant à l'origine des deux tiers de l'écart salarial entre les femmes et les hommes (INSEE, 2020). De surcroît, les femmes sont sujettes à un "plafond de verre" qui limite leur ascension vers de hauts postes (Arulampalam *et al.*, 2007). C'est dans ce contexte de ségrégation verticale que nous avons choisi d'étudier les comportements compétitifs des femmes et des hommes. L'ascension hiérarchique implique une répétition du processus compétitif, notamment dans l'accès à des promotions. Ainsi, nous nous focalisons sur les réactions après une défaite et une victoire

au moyen d'une expérience de laboratoire. Mentionner ce choix permet d'exposer le second avantage de la méthodologie expérimentale : un cadre contrôlé. Ce dernier rend notamment possible, dans notre cas, la divulgation du genre de l'opposant.e sans risquer l'interférence d'autres types de caractéristiques individuelles dans le choix de concourir ou non des sujets. Nos résultats révèlent que les hommes sur-confiants qui ont perdu face à une femme choisissent plus le tournoi et s'efforcent plus à vaincre leur adversaire, comparé à lorsqu'ils jouent face à un homologue masculin. L'illustration de ce comportement, au travers de données expérimentales, met en lumière l'effet potentiellement néfaste de la comparaison (possiblement alimentée ici par des stéréotypes) sur les carrières professionnelles des femmes. Si les hommes s'appliquent à vaincre une opposante, ils sont peut-être réticents à l'idée de voir des femmes performantes occuper des postes élevés. Les résultats de Husain *et al.* (2021) fournissent par ailleurs un exemple assez parlant. À travers un échantillon d'écoles de l'état de New York, les auteurs et autrices révèlent que les enseignants masculins sont plus susceptibles de démissionner lorsque la direction de l'école est tenue par une femme.

Dans notre premier chapitre, les abilités des individus sont donc soumis à une comparaison sociale en raison du cadre compétitif de l'étude. Les femmes ont plutôt tendance à éviter de se soumettre à la comparaison induite par le tournoi. En revanche, une fois qu'elles sont entrées en compétition, elles n'abandonnent pas nécessairement plus souvent que les hommes. Ces derniers, déjà attirés par le premier tournoi, sont plus enclins à vouloir s'imposer après une défaite contre une femme. La comparaison avec l'autre genre agissant ainsi comme un moteur de performance pour nos sujets masculins. Quels éléments peuvent expliquer ce comportement ? Selon la logique développée par Gneezy *et al.* (2009), les acquis sociaux (consolidés par un système patriarcal) constitueraient la raison principale des différences de genre en termes d'esprit de compétition. Suivant ce constat, et pour ainsi pousser plus loin l'étude de la comparaison, il serait intéressant d'étudier l'effet des normes sociales. Selon Datta Gupta *et al.* (2005), les hommes semblent être guidés dans leur choix par une sorte de convention sociale, impliquant que "les hommes doivent être compétitifs". De plus, leurs échecs sont plus souvent perçus comme un manque d'effort, comparé aux femmes (Cappelen

et al., 2019). L'analyse des comportements compétitifs en présence d'une audience, et notamment en fonction de sa composition, apparaît comme une extension pertinente du Chapitre 1. Selon Schmidt *et al.* (2021), les hommes sont plus disposés à prendre des risques en présence d'une large part de femmes. Nous pourrions envisager d'observer, dans de futurs travaux, si ces résultats sont reproductibles à l'entrée dans la compétition.

Le deuxième chapitre met en évidence les effets de la comparaison sur la perception des inégalités. L'économie vise (en autre) à étudier la répartition des richesses au sein d'une société. L'analyse des inégalités constitue donc un enjeu primordial, d'autant qu'elles ont des implications économiques sur la croissance ou le taux de criminalité (Benabou, 1996; Knell, 1999; Kelly, 2000; Stiglitz, 2015). Aujourd'hui, les 50% les plus pauvres de la population mondiale ne possède qu'à peine 2% des richesses totales contre 76% pour la moitié la plus riche (Chancel *et al.*, 2022). Ce partage inégal des ressources peut conduire à une instabilité du système social et ainsi alimenter la comparaison sociale. Dans le Chapitre 2 de cette thèse, la comparaison n'est pas appréhendée comme conséquence mais comme moteur d'inégalités. Grâce à l'environnement contrôlé permis par les expériences de laboratoire, nous distribuons différentes allocations aux sujets, créant ainsi volontairement des inégalités. Par la suite, le (ou la) joueur.euse le (ou la) plus riche a la possibilité de redistribuer équitablement les richesses. Pour tester l'influence de la comparaison sur la volonté de redistribuer, nous communiquons le niveau d'inégalités ou la décision de ne pas redistribuer dans un autre groupe. Cet autre groupe expérimente un niveau d'inégalités plus élevé. Dans ce cadre, la comparaison est supposée induire une minimisation du niveau d'inégalités perçu dans le groupe le moins inégalitaire. En effet, nos données montrent qu'observer *pire* que soi permet à l'individu le plus riche du groupe de moins redistribuer qu'en l'absence d'éléments de comparaison. En revanche, divulguer la décision de ne pas redistribuer dans le groupe le plus inégalitaire n'entraîne pas de différences significatives.

Ce travail ouvre la voie à de possibles extensions en laboratoire ou à travers un sondage. La technique d'enquête permettrait en outre d'étendre l'exploration des mécanismes de comparaison à divers autres pays. Nous pourrions ainsi imaginer des questions suivant la forme

de celles posées par l'International Social Survey Programme, dont les données ont été utilisées pour produire la Figure 2 de l'introduction de cette thèse. Les sondé.e.s pourraient être interrogé.e.s sur leur perception du niveau d'inégalités en France et différents traitements permettraient d'induire une comparaison avec d'autres pays plus ou moins inégalitaires. En l'état actuel, les groupes (et sujets) ne se distinguent que par leur niveau d'inégalités de richesse. D'autres facteurs de différenciation sont envisageables, afin de déterminer si les effets trouvés sont identiques en fonction des pairs ou de la forme que prend l'inégalité. Courant 2022, la Cour suprême étasunienne votait la déconstitutionnalisation du droit à l'avortement, entrainant une interdiction de l'IVG dans plusieurs états. Le 8 mars dernier, la France inscrivait dans la Constitution la liberté de recourir à l'avortement. L'actualité étasunienne a-t-elle contribué à cette avancée en matière de droit des femmes ? Si oui, quels facteurs en particulier ? Est-ce le retour en arrière d'une loi pourtant établie depuis 1973[26] ? Les États-Unis constituent l'une des plus grandes puissances mondiales. Dès lors, la comparaison avec tels ne fait-elle pas apparaître la menace comme plus *réaliste* aux yeux des pays développés ? Pour étudier ce phénomène, une expérience en laboratoire pourrait être conduite pour contrôler pour ces deux effets potentiels. Enfin, la comparaison peut être utilisée comme moteur de progrès (ou stagnation) mais elle peut aussi être volontairement évitée. À travers une expérience de terrain, Andreoni *et al.* (2017) ont démontré que la plupart des individus ne souhaitant pas donner à une oeuvre de charité évitent de passer devant les personnes en charge de solliciter les dons pour ne pas avoir à refuser la requête. Cet évitement permet de préserver à la fois notre image sociale (celle que l'on renvoie aux autres) et l'image de soi (celle que l'on renvoie à nous-même). Rapporté au contexte des inégalités, nous pourrions envisager de proposer aux sujets d'éviter la comparaison avec *mieux* pour leur permettre de maintenir le niveau d'inégalités dans leur groupe sans compromettre leurs images.

Le troisième chapitre de cette thèse mobilise un autre outil de la méthodologie expérimentale, à savoir les expériences de terrain. Cette méthode ne bénéficie pas d'un environnement

[26] À ce titre il convient de rappeler les mots de Simone de Beauvoir : "*il suffira d'une crise politique, économique ou religieuse pour que les droits des femmes soient remis en question*", rapportés par Monteil (2011).

aussi contrôlé que les expériences en laboratoire mais elle permet d'observer les comportements des agents en situation réelle. Leur implémentation à grande échelle offre une généralisation possible des résultats et permet ainsi de guider les décideurs publics. Notre troisième chapitre présente un intérêt économique normatif en termes de santé. Dans cette étude, nous nous intéressons aux incitations permettant d'encourager les individus à pratiquer une activité physique régulière (i.e. la marche) au moyen d'une application de santé. Les applications mobiles de santé sont de plus en plus utilisées et offrent un cadre d'étude aux chercheur.euse.s en économie. En effet, compte tenu du nombre important d'utilisateur.rice.s (i.e. 37% de la population française en 2022), l'implémentation de politiques publiques de santé à l'aide de ces plateformes permet de couvrir un large échantillon. Comme il l'a été déjà mentionné, les dépenses en termes de santé représentent un coût important. Inciter les individus à pratiquer une activité physique régulière peut s'avérer coûteux et peu efficace si les individus ne sont pas suivis. Grâce aux applications mobiles, les campagnes de sensibilisation peuvent être délivrées en continue et à moindre coût. En outre, les applications de santé ne se limitent pas à la transmission d'informations de santé. Celles-ci peuvent offrir un cadre de comparaison sociale en permettant de se connecter à des ami.e.s virtuel.le.s, des éléments de ludification (*gamification*) au moyen de badges ou de challenges ou encore des incitations monétaires. Pour notre expérience de terrain, nous avons collaborer avec l'application de suivi de pas *We Ward* qui proposent les fonctionnalités susmentionnées. De plus, l'application envoie des rappels journaliers à ses utilisateur.rice.s afin de maintenir un usage régulier de l'application et favoriser la prise de conscience du nombre de pas effectués. Différents travaux ont démontré l'efficacité des rappels dans la création de comportements vertueux : fréquentation de salle de sport (Calzolari & Nardotto, 2017), vaccination (Busso *et al.*, 2015; Chen *et al.*, 2023) ou encore l'arrêt de la cigarette (Rodgers *et al.*, 2005). Dans notre expérience, nous cherchons à identifier les effets de plusieurs types de rappels sur la fréquence d'utilisation de l'application et sur les habitudes de marche.

Nous conduisons ainsi un Essai Randomisé Controllé (ERC) auprès de plus de 20 000 individus pour tester l'effet de nouveaux rappels. Ceux-ci mettent en avant un aspect com-

portemental précis : les coûts irrécupérables, la comparaison avec les pairs ou avec ses propres performances. Nos résultats montrent que la fréquence d'utilisation de *WeWard* augmente avec des messages poussant à la comparaison, qu'elle soit en rapport avec ses propres performances ou celles de pairs. Plus spécifiquement, les individus ont tendance à être plus enclins à ouvrir l'application lorsqu'ils recoivent un rappel mettant en avant la performance de pairs. Concernant la récurrence, c'est la comparaison avec ses performances passées qui incite les utilisateur.rice.s à ouvrir plus régulièrement l'application durant la journée. En revanche, aucun des rappels implémentés n'a permis d'inciter les individus à marcher plus. Les habitudes de marche sont essentiellement influencées par des facteurs externes, tels que les jours de la semaine ou les fêtes de fin d'année. Ce constat suggère que les comportements de santé, et notamment ceux associés à une activité aussi commune que la marche, sont difficiles à modifier. Nos observations concernant l'utilisation de l'application invitent quant à elles à continuer d'exploiter cet outil. Les habitudes de marche n'ont pas été significativement augmentées mais certains rappels envoyés via les applications de santé peuvent attirer l'attention des utilisateur.rice.s. Nos nouveaux rappels, adaptés à différentes applications, pourraient être utilisés pour d'autres types de comportements de santé comme la perte de poids par exemple. Avant de mener un tel travail, il conviendra d'appréhender l'influence des caractéristiques individuelles (tel que le genre) dans l'utilisation des applications ou la pratique d'une activité physique.

En effet, des travaux ont montré que les hommes sont généralement motivés par des facteurs internes (tels que la compétition ou le challenge) alors que les femmes sont plutôt stimulées par leur apparence physique par exemple (Kilpatrick *et al.*, 2005; Azevedo *et al.*, 2007; Egli *et al.*, 2011; Lauderdale *et al.*, 2015; Molanorouzi *et al.*, 2015). En outre, la pratique d'une activité physique tend à satisfaire un objectif de minceur chez les femmes et de développement de la musculature chez les hommes (Cockburn & Clarke, 2002; Allender *et al.*, 2006). Ces différences pourraient expliquer la sous-représentation des femmes dans certaines pratiques, comme les marathons (dont les femmes ont été historiquement exclues). En 2021, les femmes ne représentaient que 20% des marathonien.ne.s contre 40% dans les

courses de 10km[27]. Ces disparités peuvent également conduire à des différences d'utilisation d'application de santé. Mobilisant les données issues du Chapitre 3, nous questionnons l'effet des incitations monétaires proposées par les applications de santé dans un quatrième chapitre. Nos résultats complémentent ceux du troisième chapitre puisqu'ils poussent la réflexion sur l'efficacité des applications de santé, en fonction des caractéristiques des utilisateur.rice.s. Pour ce faire, nous mobilisons un autre outil économétrique, adapté à cette nouvelle question de recherche. Au moyen d'une régression sur discontinuité (RDD), nous isolons l'effet du passage d'un seuil de récompense sur les pratiques d'utilisation de l'application *We Ward*.

Nos données révèlent que les hommes et les femmes ne diffèrent pas significativement dans leur pratique d'utilisation en fonction des récompenses proposées par l'application. Atteindre un seuil de récompense augmente les interactions avec *We Ward* de manière équivalente pour les utilisateurs et utilisatrices. Ce résultat tend à affirmer l'efficacité de la fonctionnalité principale de *We Ward*, incitant les usager.ère.s à se connecter à l'application. En revanche, nos données mettent en avant un biais qu'il convient d'adresser lors de la récolte de données expérimentales. Le travail mené dans ce quatrième chapitre a tout d'abord été motivé par la découverte de différences de genre dans les pratiques de marche. Les femmes apparaissent marcher significativement moins que les hommes alors qu'elles ouvrent pourtant l'application plus régulièrement. Face à ce constat, nous avons décidé de recourir à des données administratives pour comparer nos donnés enregistrées via un podomètre à des données déclaratives. Ces dernières illustrent plutôt un écart de genre en faveur des femmes, qui déclarent marcher plus. Dans la mesure où nos données *We Ward* dépendent de l'ouverture de l'application par l'utilisateur.rice, celles-ci sont soumises à un problème d'attrition. Comme évoqué au cours du quatrième chapitre, l'attrition peut conduire à une sur-estimation du nombre de pas effectués par les hommes. Si les utilisatrices ouvrent plus souvent l'application, elles enregistrent mécaniquement plus souvent des petits nombres de pas, comparé aux hommes. Une piste qui pourrait être explorée pour expliquer cette tendance réside, là encore, dans la comparaison

[27]"Marathon de Paris : Pourquoi les femmes participent moins au marathon que les hommes ?", 20 minutes (2024): *https : //www.20minutes.fr/sport/marathon_de_paris/4072623 − 20240309 − marathon − paris − pourquoi − femmes − participent − moins − marathon − hommes*

avec les pairs. Le nombre de pas validés sur l'application n'est visible auprès des autres utilisateur.rice.s qu'à condition que l'usager.ère valide ses pas. Dès lors, un individu pourrait avoir intérêt à ne pas valider lorsque son nombre de pas n'égalise (ou ne surpasse) pas celui de ses pairs. Malheureusement, ne disposant pas d'informations sur le nombre d'ami.e.s virtuel.le.s que possèdent nos sujets, nous ne pouvons creuser cette piste. Il serait intéressant d'explorer cette voie, faisant ainsi écho à notre premier chapitre sur l'esprit de compétition et l'impact des pairs sur les performances, ici en termes de santé.

Ainsi s'achève ce travail de thèse portant sur l'impact de la comparaison sur les performances, opinions et comportements de santé. Le premier chapitre démontre que la comparaison peut engendrer (chez les hommes) une volonté de surpasser une adversaire lorsque les résultats ne sont pas en adéquation avec les attentes. Dans le deuxième chapitre, la comparaison agit comme un obstacle dans la poursuite d'un objectif d'égalité. Enfin, se comparer avec des pairs peut permettre une plus grande interaction avec des applications de santé, et *in fine*, l'adoption de comportements plus vertueux. Nos résultats ainsi que l'apprentissage des différentes méthodologies qui y sont associées invitent à poursuivre la réflexion sur l'effet de la comparaison, notamment à travers les extensions proposées.

"*Les progrès réalisés doivent être utilisés pour nourrir les progrès suivants, et non pour alimenter l'autocongratulation*" (p.224, Piketty, 2021).

Bibliography

Abel, Andrew B. 1990. *Asset prices under habit formation and catching up with the Joneses*.

Acland, Dan, & Levy, Matthew R. 2015. Naiveté, projection bias, and habit formation in gym attendance. *Management Science*, **61**(1), 146–160.

Ainsworth, Barbara E. 2000. Issues in the assessment of physical activity in women. *Research Quarterly for Exercise and Sport*, **71**(sup2), 37–42.

Alan, Sule, Ertac, Seda, Kubilay, Elif, & Loranth, Gyongyi. 2020. Understanding gender differences in leadership. *The Economic Journal*, **130**(626), 263–289.

Albrecht, Konstanze, Von Essen, Emma, Parys, Juliane, & Szech, Nora. 2013. Updating, self-confidence, and discrimination. *European Economic Review*, **60**, 144–169.

Alderson, Arthur S, & Katz-Gerro, Tally. 2016. Compared to whom? Inequality, social comparison, and happiness in the United States. *Social Forces*, **95**(1), 25–54.

Alesina, Alberto, & Angeletos, George-Marios. 2005. Fairness and redistribution. *American Economic Review*, **95**(4), 960–980.

Alesina, Alberto, & La Ferrara, Eliana. 2000. Participation in heterogeneous communities. *The Quarterly Journal of Economics*, **115**(3), 847–904.

Alesina, Alberto, Di Tella, Rafael, & MacCulloch, Robert. 2004. Inequality and happiness: Are Europeans and Americans different? *Journal of Public Economics*, **88**(9-10), 2009–2042.

Alexander, Gwen L, Divine, George W, Couper, Mick P, McClure, Jennifer B, Stopponi, Melanie A, Fortman, Kristine K, Tolsma, Dennis D, Strecher, Victor J, & Johnson, Christine Cole. 2008. Effect of incentives and mailing features on online health program enrollment. *American Journal of Preventive Medicine*, **34**(5), 382–388.

Allender, Steven, Cowburn, Gill, & Foster, Charlie. 2006. Understanding participation in sport and physical activity among children and adults: A review of qualitative studies. *Health Education Research*, **21**(6), 826–835.

Almas, I, Cappelen, Alexander W, Sorensen, EO, & Tungodden, Bertil. 2022. Attitudes to Inequality: Preferences and Beliefs. *IFS Deaton Review of Inequalities*.

Almås, Ingvild, Cappelen, Alexander W, Sørensen, Erik Ø, & Tungodden, Bertil. 2010. Fairness and the development of inequality acceptance. *Science*, **328**(5982), 1176–1178.

Almås, Ingvild, Cappelen, Alexander W, Salvanes, Kjell G, Sørensen, Erik Ø, & Tungodden, Bertil. 2016. Willingness to compete: Family matters. *Management Science*, **62**(8), 2149–2162.

Almas, Ingvild, Cappelen, Alexander W, Sorensen, EO, & Tungodden, Bertil. 2021. Fairness across the World: Preferences and Beliefs. *Federal Reserve Bank of New York*.

Amanatullah, Emily T, & Morris, Michael W. 2010. Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology*, **98**(2), 256.

Ambady, Nalini, Shih, Margaret, Kim, Amy, & Pittinsky, Todd L. 2001. Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, **12**(5), 385–390.

Andersen, Ross E. 1999. Exercise, an active lifestyle, and obesity: Making the exercise prescription work. *The Physician and Sports Medicine*, **27**(10), 41–50.

Andreoni, James, & Miller, John. 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, **70**(2), 737–753.

Andreoni, James, Rao, Justin M, & Trachtman, Hannah. 2017. Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy*, **125**(3), 625–653.

Apicella, Coren L, Demiral, Elif E, & Mollerstrom, Johanna. 2017. No gender difference in willingness to compete when competing against self. *American Economic Review*, **107**(5), 136–140.

Argouarc'h, Julie, & Calavrezo, Oana. 2013. La répartition des hommes et des femmes par métiers: une baisse de la ségrégation depuis 30 ans. *Dares analyses*, **79**.

Aronson, Joshua, Lustina, Michael J, Good, Catherine, Keough, Kelli, Steele, Claude M, & Brown, Joseph. 1999. When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, **35**(1), 29–46.

Arrow, Kenneth J. 1972. Some mathematical models of race discrimination in the labor market. *Racial discrimination in economic life*, 187–204.

Arsenio, William F, & Willems, Chris. 2017. Adolescents' conceptions of national wealth distribution: Connections with perceived societal fairness and academic plans. *Developmental Psychology*, **53**(3), 463.

Arulampalam, Wiji, Booth, Alison L, & Bryan, Mark L. 2007. Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *Industrial and Labor Relations Review*, **60**(2), 163–186.

Asher, Maya, Asnaani, Anu, & Aderka, Idan M. 2017. Gender differences in social anxiety disorder: A review. *Clinical Psychology Review*, **56**, 1–12.

Azevedo, Mario Renato, Araújo, Cora Luiza Pavin, Reichert, Felipe Fossati, Siqueira, Fernando Vinholes, da Silva, Marcelo Cozzensa, & Hallal, Pedro Curi. 2007. Gender differences in leisure-time physical activity. *International Journal of Public Health*, **52**, 8–15.

Azmat, Ghazala, & Petrongolo, Barbara. 2014. Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, **30**, 32–40.

Bagein, Guillaume, Costemalle, Vianney, Deroyon, Thomas, Hazo, Jean-Baptiste, Naouri, Diane, Pesonel, Elise, & Vilain, Annick. 2022. L'état de santé de la population en France. *Dossiers de la DREES*.

Baier, Alexandra, Davis, Brent J, Jaber-Lopez, Tarek, & Seidl, Michael. 2018. *Gender, competition and the effect of feedback and task: An experiment (No. 062)*. Tech. rept. *Working Paper Forschungsförderung*.

Baier, Alexandra, Davis, Brent, & Jaber-Lopez, Tarek. 2024. Gender, choice of task, and the effect of feedback on competition: An experiment. *Journal of Economic Psychology*, 102743.

Baker, CW, & Brownell, KD. 2000. Physical activity and maintenance of weight loss: Physiological and psychological mechanisms.

Balafoutas, Loukas, & Sutter, Matthias. 2012. Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, **335**(6068), 579–582.

Bandura, Albert, & Walters, Richard H. 1963. Social learning and personality development.

Barnett, Richard C, Bhattacharya, Joydeep, & Bunzel, Helle. 2010. Choosing to keep up with the Joneses and income inequality. *Economic Theory*, **45**, 469–496.

Bassett, David R. 2000. Validity of four motion sensors in measuring moderate intensity physical activity. *Medicine & Science in Sports & Exercise*, **32**(9), S471–S480.

Bateman, Angus J. 1948. Intra-sexual selection in Drosophila. *Heredity*, **2**(3), 349–368.

Bavetta, Sebastiano, Li Donni, Paolo, & Marino, Maria. 2019. An empirical analysis of the determinants of perceived inequality. *Review of Income and Wealth*, **65**(2), 264–292.

Baxter, Amanda J, Scott, Kate M, Ferrari, Alize J, Norman, Rosana E, Vos, Theo, & Whiteford, Harvey A. 2014. Challenging the myth of an "epidemic" of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depression and Anxiety*, **31**(6), 506–516.

Beaman, Lori, Duflo, Esther, Pande, Rohini, & Topalova, Petia. 2012. Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, **335**(6068), 582–586.

Becker, Gary S. 1957. *The economics of discrimination*. University of Chicago press.

Benabou, Roland. 1996. Inequality and growth. *NBER macroeconomics annual*, **11**, 11–74.

Benabou, Roland, & Ok, Efe A. 2001. Social mobility and the demand for redistribution: The POUM hypothesis. *The Quarterly Journal of Economics*, **116**(2), 447–487.

Berlin, Noémi, & Dargnies, Marie-Pierre. 2016. Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, **130**, 320–336.

Bertrand, Marianne, & Mullainathan, Sendhil. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, **94**(4), 991–1013.

Bicchieri, Cristina, & Xiao, Erte. 2009. Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, **22**(2), 191–208.

Binmore, Ken. 2010. Social norms or social preferences? *Mind & Society*, **9**, 139–157.

Blau, Francine D, & Kahn, Lawrence M. 2017. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, **55**(3), 789–865.

Blommaert, Lieselotte, Coenders, Marcel, & Van Tubergen, Frank. 2014. Discrimination of Arabic-named applicants in the Netherlands: An internet-based field experiment examining different phases in online recruitment procedures. *Social forces*, **92**(3), 957–982.

Bohnet, Iris, & Zeckhauser, Richard. 2004. Social comparisons in ultimatum bargaining. *The Scandinavian Journal of Economics*, **106**(3), 495–510.

Bolton, Gary E, & Ockenfels, Axel. 2000. ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, **91**(1), 166–193.

Booth, Alison, & Nolen, Patrick. 2012. Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, **81**(2), 542–555.

Bordalo, Pedro, Coffman, Katherine, Gennaioli, Nicola, & Shleifer, Andrei. 2019. Beliefs about gender. *American Economic Review*, **109**(3), 739–73.

Bowles, Hannah Riley, Babcock, Linda, & Lai, Lei. 2007. Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and human decision Processes*, **103**(1), 84–103.

Broocks, Andreas, Bandelow, Borwin, Pekrun, Gunda, George, Annette, Meyer, Tim, Bartmann, Uwe, Hillmer-Vogel, Ursula, & Rüther, Eckart. 1998. Comparison of aerobic exercise, clomipramine, and placebo in the treatment of panic disorder. *American Journal of Psychiatry*, **155**(5), 603–609.

Brown, Wendy, Ringuet, Caroline, Trost, Stewart G, & Jenkins, David. 2001. Measurement of energy expenditure of daily tasks among mothers of young children. *Journal of Science and Medicine in Sport*, **4**(4), 379–385.

Bueno-Antequera, Javier, & Munguia-Izquierdo, Diego. 2020. Exercise and depressive disorder. *Physical Exercise for Human Health*, 271–287.

Burleigh, Tyler J, & Meegan, Daniel V. 2013. Keeping up with the Joneses affects perceptions of distributive justice. *Social Justice Research*, **26**, 120–131.

Burow, Norma, Beblo, Miriam, Beninger, Denis, & Schröder, Melanie. 2017. Why Do Women Favor Same-Gender Competition. *Evidence from a choice experiment. DIW Discussion Papers*, **1662**.

Buser, Thomas. 2016. The impact of losing in a competition on the willingness to seek further challenges. *Management Science*, **62**(12), 3439–3449.

Buser, Thomas, & Yuan, Huaiping. 2019. Do women give up competing more easily? Evidence from the lab and the Dutch math olympiad. *American Economic Journal: Applied Economics*, **11**(3), 225–252.

Buser, Thomas, Gerhards, Leonie, & Van Der Weele, Joël. 2018. Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, **56**, 165–192.

Buser, Thomas, van den Assem, Martijn J, & van Dolder, Dennie. 2023. Gender and willingness to compete for high stakes. *Journal of Economic Behavior & Organization*, **206**, 350–370.

Busso, Matías, Cristia, Julian, & Humpage, Sarah. 2015. Did you get your shots? Experimental evidence on the role of reminders. *Journal of Health Economics*, **44**, 226–237.

Cai, Xiqian, Lu, Yi, Pan, Jessica, & Zhong, Songfa. 2019. Gender gap under pressure: Evidence from China's National College entrance examination. *Review of Economics and Statistics*, **101**(2), 249–263.

Calzolari, Giacomo, & Nardotto, Mattia. 2017. Effective reminders. *Management Science*, **63**(9), 2915–2932.

Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Cappelen, Alexander W, Falch, Ranveig, & Tungodden, Bertil. 2019. The boy crisis: Experimental evidence on the acceptance of males falling behind. *NHH Dept. of Economics Discussion Paper*.

Carrera, Mariana, Royer, Heather, Stehr, Mark, & Sydnor, Justin. 2018. Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members. *Journal of Health Economics*, **58**, 202–214.

Cason, Timothy N, & Mui, Vai-Lam. 1998. Social influence in the sequential dictator game. *Journal of Mathematical Psychology*, **42**(2-3), 248–265.

Cason, Timothy N, Masters, William A, & Sheremeta, Roman M. 2010. Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, **94**(9-10), 604–611.

Castleman, Benjamin L, & Page, Lindsay C. 2015. Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, **115**, 144–160.

Cattaneo, Matias D, Titiunik, Rocío, Vazquez-Bare, Gonzalo, & Keele, Luke. 2016. Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics*, **78**(4), 1229–1248.

Centola, Damon. 2013. Social media and the science of health behavior. *Circulation*, **127**(21), 2135–2144.

Chancel, Lucas, Piketty, Thomas, Saez, Emmanuel, & Zucman, Gabriel. 2022. *World Inequality Report 2022*. Harvard University Press.

Charness, Gary, & Gneezy, Uri. 2009. Incentives to exercise. *Econometrica*, **77**(3), 909–931.

Chen, Daniel L, Schonger, Martin, & Wickens, Chris. 2016. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, **9**, 88–97.

Chen, Nuole, Trump, Kris-Stella, Hall, Stacy, & Le, Quan. 2023. The effect of postcard reminders on vaccinations among the elderly: A block-randomized experiment. *Behavioural Public Policy*, **7**(2), 240–265.

Cheung, Felix, & Lucas, Richard E. 2016. Income inequality is associated with stronger social comparison effects: The effect of relative income on life satisfaction. *Journal of Personality and Social Psychology*, **110**(2), 332.

Chou, Hui-Tzu Grace, & Edge, Nicholas. 2012. "They are happier and having better lives than I am": the impact of using Facebook on perceptions of others' lives. *Cyberpsychology, Behavior, and Social networking*, **15**(2), 117–121.

Choudhary, Vivek, Shunko, Masha, Netessine, Serguei, & Koo, Seongjoon. 2022. Nudging drivers to safety: Evidence from a field experiment. *Management Science*, **68**(6), 4196–4214.

Christakis, Nicholas A, & Fowler, James H. 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, **357**(4), 370–379.

Chung, Y Barry. 2001. Work discrimination and coping strategies: Conceptual frameworks for counseling lesbian, gay, and bisexual clients. *The Career Development Quarterly*, **50**(1), 33–44.

Clark, Andrew E, & d'Ambrosio, Conchita. 2015. Attitudes to income inequality: Experimental and survey evidence. *Pages 1147–1208 of: Handbook of income distribution*, vol. 2. Elsevier.

Clark, Andrew E, & Oswald, Andrew J. 1996. Satisfaction and comparison income. *Journal of Public Economics*, **61**(3), 359–381.

Clark, Andrew E, & Senik, Claudia. 2010. Who compares to whom? The anatomy of income comparisons in Europe. *The Economic Journal*, **120**(544), 573–594.

Clark, Andrew E, Frijters, Paul, & Shields, Michael A. 2008. Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic Literature*, **46**(1), 95–144.

Cockburn, Claudia, & Clarke, Gill. 2002. "Everybody's looking at you!": Girls negotiating the "femininity deficit" they incur in physical education. *Pages 651–665 of: Women's Studies International Forum*, vol. 25. Elsevier.

Collins, Brianna S, Marshall, Alison L, & Miller, Yvette D. 2007. Physical activity in women with young children: How can we assess "anything that's not sitting"? *Women & Health*, **45**(2), 95–116.

Cortes, Patricia, & Pan, Jessica. 2018. Occupation and gender. *The Oxford handbook of women and the economy*, 425–452.

Costa, Dora L, & Kahn, Matthew E. 2013. Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, **11**(3), 680–702.

Coutts, Alexander. 2019. Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, **22**(2), 369–395.

Crosetto, Paolo, & Filippin, Antonio. 2013. The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*, **47**(1), 31–65.

Cruces, Guillermo, Perez-Truglia, Ricardo, & Tetaz, Martin. 2013. Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, **98**, 100–112.

Dana, Jason, Weber, Roberto A, & Kuang, Jason Xi. 2007. Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, **33**, 67–80.

Darwin, Charles. 1871. *The descent of man, and selection in relation to sex*. Princeton University Press.

Datta Gupta, Nabanita, Poulsen, Anders, & Villeval, Marie Claire. 2005. *Male and female*

competitive behavior-experimental evidence. Tech. rept. *GATE Working Paper No. WP 05-12*.

Datta Gupta, Nabanita, Poulsen, Anders, & Villeval, Marie Claire. 2013. Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, **51**(1), 816–835.

Davidai, Shai, & Gilovich, Thomas. 2015. Building a more mobile America—One income quintile at a time. *Perspectives on Psychological Science*, **10**(1), 60–71.

De La Torre, Steven, Spruijt-Metz, Donna, Farias, Albert J, *et al.* 2021. Associations among wearable activity tracker use, exercise motivation, and physical activity in a cohort of cancer survivors: Secondary data analysis of the health information national trends survey. *JMIR cancer*, **7**(2), e24828.

De Paola, Maria, & Scoppa, Vincenzo. 2015. Gender discrimination and evaluators' gender: Evidence from Italian academia. *Economica*, **82**(325), 162–188.

Denny, Elizabeth B, & Hunt, R Reed. 1992. Affective valence and memory in depression: Dissociation of recall and fragment completion. *Journal of Abnormal Psychology*, **101**(3), 575.

Desine, Stacy, Master, Hiral, Annis, Jeffrey, Hughes, Andrew, Roden, Dan M, Harris, Paul A, & Brittain, Evan L. 2023. Daily step counts before and after the COVID-19 pandemic among all of US research participants. *JAMA Network Open*, **6**(3), e233526–e233526.

DiBenedetto, Barbara, & Tittle, Carol Kehr. 1990. Gender and adult roles: Role commitment of women and men in a job-family trade-off context. *Journal of Counseling Psychology*, **37**(1), 41–48.

Dilmaghani, Maryam. 2022. Chess girls don't cry: Gender composition of games and effort in competitions among the super-elite. *Journal of Economic Psychology*, **89**, 102482.

Dreber, Anna, Von Essen, Emma, & Ranehill, Eva. 2014. Gender and competition in adolescence: Task matters. *Experimental Economics*, **17**(1), 154–172.

DREES. 2022. *Les dépenses de santé en 2021.* Tech. rept.

Du, Hongfei, & King, Ronnel B. 2022. What predicts perceived economic inequality? The roles of actual inequality, system justification, and fairness considerations. *British Journal of Social Psychology*, **61**(1), 19–36.

Duffy, John, & Feltovich, Nick. 1999. Does observation of others affect learning in strategic environments? An experimental study. *International Journal of Game theory*, **28**, 131–152.

Easterlin, Richard A. 1974. Does economic growth improve the human lot? Some empirical evidence. *Pages 89–125 of: Nations and households in economic growth.* Elsevier.

Edo, Anthony, & Jacquemet, Nicolas. 2014. Discrimination à l'embauche selon l'origine et le genre: défiance indifférenciée ou ciblée sur certains groupes? *Economie et Statistique/Economics and Statistics*, 155–172.

Egli, Trevor, Bland, Helen W, Melton, Bridget F, & Czech, Daniel R. 2011. Influence of age, sex, and race on college students' exercise motivation of physical activity. *Journal of American College Health*, **59**(5), 399–406.

Ellingrud, Kweilin, Madgavkar, Anu, & Manyika, James. 2016. *The power of parity: Advancing women's equality in the United States.* Tech. rept. McKinsey Global Institute.

Ellis, Jessica, Fosdick, Bailey K, & Rasmussen, Chris. 2016. Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PloS one*, **11**(7), e0157447.

Engel, Christoph. 2011. Dictator games: A meta study. *Experimental Economics*, **14**, 583–610.

Ertac, Seda, & Szentes, Balazs. 2011. *The effect of information on gender differences in competitiveness: Experimental evidence.* Tech. rept. Working paper.

European Commission. 2021. *Gender pay gap statistics.* Tech. rept.

Faizah, Aaya, Hardian, Alifah Fatimah Azzahra, Nandini, Rania Devina, Handayani, Putu Wuri, & Harahap, Nabila Cyldea. 2024. The Influence of Incentive-Based Mobile Fitness Apps on Users' Continuance Intention With Gender Moderation Effects: Quantitative and Qualitative Study. *JMIR Human Factors*, **11**, e50957.

Falk, Armin, & Knell, Markus. 2004. Choosing the Joneses: Endogenous goals and reference standards. *The Scandinavian Journal of Economics*, **106**(3), 417–435.

Fang, Chao, Zhang, Ernest, & Zhang, Junfu. 2021. Do women give up competing more easily? Evidence from speedcubers. *Economics Letters*, **205**, 109943.

Fehr, Ernst, & Schmidt, Klaus M. 1999. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, **114**(3), 817–868.

Feng, Wenting, Tu, Rungting, & Hsieh, Peishan. 2020. Can gamification increases consumers' engagement in fitness apps? The moderating role of commensurability of the game elements. *Journal of Retailing and Consumer Services*, **57**, 102229.

Ferrer-i-Carbonell, Ada. 2005. Income and well-being: An empirical analysis of the comparison income effect. *Journal of Public Economics*, **89**(5-6), 997–1019.

Festinger, Leon. 1954. A theory of social comparison processes. *Human Relations*, **7**(2), 117–140.

Fischer, Stefanie. 2017. The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics*, **46**, 211–226.

Fitzgerald, Amanda, Fitzgerald, Noelle, & Aherne, Cian. 2012. Do peers matter? A review of peer and/or friends' influence on physical activity among American adolescents. *Journal of Adolescence*, **35**(4), 941–958.

Fjeldsoe, Brianna, Marshall, Alison, & Miller, Yvette. 2009. Measurement properties of the Australian women's activity survey. *Medicine and Science in Sports and Exercise*, **41**(5), 1020–1033.

Fjeldsoe, Brianna S, Miller, Yvette D, & Marshall, Alison L. 2010. MobileMums: A randomized controlled trial of an SMS-based physical activity intervention. *Annals of behavioral medicine*, **39**(2), 101–111.

Forsythe, Robert, Horowitz, Joel L, Savin, Nathan E, & Sefton, Martin. 1994. Fairness in simple bargaining experiments. *Games and Economic Behavior*, **6**(3), 347–369.

Franken, Rob, Bekhuis, Hidde, & Tolsma, Jochem. 2023. Kudos make you run! How runners influence each other on the online social network Strava. *Social Networks*, **72**, 151–164.

Franzoi, Stephen L, Vasquez, Kris, Sparapani, Erin, Frost, Katherine, Martin, Jessica, & Aebly, Megan. 2012. Exploring body comparison tendencies: Women are self-critical whereas men are self-hopeful. *Psychology of Women Quarterly*, **36**(1), 99–109.

Fredrickson, Barbara L, & Roberts, Tomi-Ann. 1997. Objectification theory: Toward understanding women's lived experiences and mental health risks. *Psychology of Women Quarterly*, **21**(2), 173–206.

Free, Caroline, Whittaker, R, Knight, R, Abramsky, T, Rodgers, A, & Roberts, IG. 2009. Txt2stop: A pilot randomised controlled trial of mobile phone-based smoking cessation support. *Tobacco Control*, **18**(2), 88–91.

Freyne, Jill, Yin, Jie, Brindal, Emily, Hendrie, Gilly A, Berkovsky, Shlomo, & Noakes, Manny. 2017. Push notifications in diet apps: influencing engagement times and tasks. *International Journal of Human–Computer Interaction*, **33**(10), 833–845.

Gali, Jordi. 1994. Keeping up with the Joneses: Consumption externalities, portfolio choice, and asset prices. *Journal of Money, Credit and Banking*, **26**(1), 1–8.

García-Cueto, Eduardo, Rodríguez-Díaz, Francisco Javier, Bringas-Molleda, Carolina, López-Cepero, Javier, Paíno-Quesada, Susana, & Rodríguez-Franco, Luis. 2015. Development of the gender role attitudes scale (GRAS) amongst young Spanish people. *International Journal of Clinical and Health Psychology*, **15**(1), 61–68.

Geraldes, Diogo. 2020. Women dislike competing against men. *Available at SSRN 3741649*.

Gill, David, & Prowse, Victoria. 2012. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, **102**(1), 469–503.

Gill, David, & Prowse, Victoria. 2014. Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, **5**(2), 351–376.

Gillen, Ben, Snowberg, Erik, & Yariv, Leeat. 2019. Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, **127**(4), 1826–1863.

Gimpelson, Vladimir, & Treisman, Daniel. 2018. Misperceiving inequality. *Economics & Politics*, **30**(1), 27–54.

Gneezy, Uri, & Rustichini, Aldo. 2004. Gender and competition at a young age. *American Economic Review*, **94**(2), 377–381.

Gneezy, Uri, Niederle, Muriel, & Rustichini, Aldo. 2003. Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, **118**(3), 1049–1074.

Gneezy, Uri, Leonard, Kenneth L, & List, John A. 2009. Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, **77**(5), 1637–1664.

Goel, Rahul, Oyebode, Oyinlola, Foley, Louise, Tatah, Lambed, Millett, Christopher, & Woodcock, James. 2023. Gender differences in active travel in major cities across the world. *Transportation*, **50**(2), 733–749.

Goeschl, Timo, Kettner, Sara Elisa, Lohse, Johannes, & Schwieren, Christiane. 2018. From social information to social norms: Evidence from two experiments on donation behaviour. *Games*, **9**(4), 91.

Golbus, Jessica R, Shi, Jieru, Gupta, Kashvi, Stevens, Rachel, Jeganathan, V Swetha E, Luff, Evan, Boyden, Thomas, Mukherjee, Bhramar, Kohnstamm, Sarah, Taralunga, Vlad,

et al. 2024. Text Messages to Promote Physical Activity in Patients With Cardiovascular Disease: A Micro-Randomized Trial of a Just-In-Time Adaptive Intervention. *Circulation: Cardiovascular Quality and Outcomes*, e010731.

Goldin, Claudia. 2014. A grand gender convergence: Its last chapter. *American Economic Review*, **104**(4), 1091–1119.

Goldin, Claudia. 2015. Gender and the undergraduate economics major: Notes on the undergraduate economics major at a highly selective liberal arts college. *Manuscript, April*, **12**.

Groshen, Erica L. 1991. The structure of the female/male wage differential: Is it who you are, what you do, or where you work? *Journal of Human Resources*, 457–472.

Guiso, Luigi, Monte, Ferdinando, Sapienza, Paola, & Zingales, Luigi. 2008. Culture, gender, and math. *Science*, **320**(5880), 1164–1165.

Güth, Werner, Schmittberger, Rolf, & Schwarze, Bernd. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, **3**(4), 367–388.

Habla, Wolfgang, & Muller, Paul. 2021. Experimental evidence of limited attention at the gym. *Experimental Economics*, **24**(4), 1156–1184.

Halpern, David. 2015. *Inside the nudge unit: How small changes can make a big difference*. Random House.

Hardeman, Wendy, Houghton, Julie, Lane, Kathleen, Jones, Andy, & Naughton, Felix. 2019. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, **16**, 1–21.

Hauge, Karen Evelyn, Kotsadam, Andreas, & Riege, Anine. 2023. Culture and gender differences in willingness to compete. *The Economic Journal*, **133**(654), 2403–2426.

Hauser, Oliver P, & Norton, Michael I. 2017. (Mis)perceptions of inequality. *Current Opinion in Psychology*, **18**, 21–25.

Haut conseil à l'égalité entre les femmes et les hommes. 2023. *Rapport annuel 2023 sur l'état des lieux du sexisme en France*.

He, Shuya, & Noussair, Charles. 2021. *Don't tell anyone I lost to a girl! Gender stereotypes and hiding low performance*. Tech. rept. *Mimeo*.

He, Shuya, & Noussair, Charles N. 2024. Gender stereotypes and hiding low performance. *Economic Inquiry*, **62**(2), 525–542.

Head, Katharine J, Noar, Seth M, Iannarino, Nicholas T, & Harrington, Nancy Grant. 2013. Efficacy of text messaging-based interventions for health promotion: A meta-analysis. *Social Science & Medicine*, **97**, 41–48.

Heatherington, Laurie, Daubman, Kimberly A, Bates, Cynthia, Ahn, Alicia, Brown, Heather, & Preston, Camille. 1993. Two investigations of "female modesty" in achievement situations. *Sex Roles*, **29**(11), 739–754.

Hersch, Joni. 2007. *Sex discrimination in the labor market*. Now Publishers Inc USA.

Heslin, Peter A, Bell, Myrtle P, & Fletcher, Pinar O. 2012. The devil without and within: A conceptual model of social cognitive processes whereby discrimination leads stigmatized minorities to become discouraged workers. *Journal of Organizational Behavior*, **33**(6), 840–862.

Holzmeister, Felix, & Pfurtscheller, Armin. 2016. oTree: The "bomb" risk elicitation task. *Journal of Behavioral and Experimental Finance*, **10**, 105–108.

Hong, Fuhai, Huang, Wei, & Zhao, Xiaojian. 2019. Sunk cost as a self-management device. *Management Science*, **65**(5), 2216–2230.

Hoy, Christopher, & Mager, Franziska. 2021. Why are relatively poor people not more supportive of redistribution? Evidence from a randomized survey experiment across ten countries. *American Economic Journal: Economic Policy*, **13**(4), 299–328.

Husain, Aliza N, Matsa, David A, & Miller, Amalia R. 2021. Do male workers prefer male leaders? An analysis of principals' effects on teacher retention. *Journal of Human Resources*, 1118–9838R2.

Hvidberg, Kristoffer B, Kreiner, Claus, & Stantcheva, Stefanie. 2020. *Social positions and fairness views on inequality.* Tech. rept. National Bureau of Economic Research.

INJEP, & MEDES. 2020. *Enquête nationale sur les pratiques physiques et sportives (ENPPS).* Tech. rept. PROGEDO-ADISP.

INJEP, CREDOC, & DS. 2020. *Baromètre national des pratiques sportives.* Tech. rept. PROGEDO-ADISP.

Inoue, Kosuke, Tsugawa, Yusuke, Mayeda, Elizabeth Rose, & Ritz, Beate. 2023. Association of daily step patterns with mortality in US adults. *JAMA Network Open*, **6**(3), e235174–e235174.

INSEE. 2020. *Écarts de rémunération femmes-hommes : surtout l'effet du temps de travail et de l'emploi occupé.* Tech. rept. 1803.

ISCA & CEBR, International Sport and Culture Association (ISCA), Centre for Economics and Business Research (CEBR). 2015. *The economic cost of physical inactivity in Europe.* Tech. rept.

Jestico, Ben, Nelson, Trisalyn, & Winters, Meghan. 2016. Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, **52**, 90–97.

Joinson, Adam N. 2008. Looking at, looking up or keeping up with people? Motives and use of Facebook. *Pages 1027–1036 of: Proceedings of the SIGCHI conference on Human Factors in Computing Systems.*

Kahan, Dan M, Jenkins-Smith, Hank, & Braman, Donald. 2011. Cultural cognition of scientific consensus. *Journal of Risk Research*, **14**(2), 147–174.

Kamas, Linda, & Preston, Anne. 2012. The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, **83**(1), 82–97.

Karadja, Mounir, Mollerstrom, Johanna, & Seim, David. 2017. Richer (and holier) than thou? The effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics*, **99**(2), 201–212.

Katz, Sandra, Allbritton, David, Aronis, John, Wilson, Christine, & Soffa, Mary Lou. 2006. Gender, achievement, and persistence in an undergraduate computer science program. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, **37**(4), 42–57.

Kelley, Jonathan, & Zagorski, Krzysztof. 2004. Economic change and the legitimation of inequality: The transition from socialism to the free market in Central-East Europe. *Research in Social Stratification and Mobility*, **22**, 319–364.

Kelly, Morgan. 2000. Inequality and crime. *Review of Economics and Statistics*, **82**(4), 530–539.

Kilmann, Ralph. 1983. Thomas-Kilmann conflict mode instrument. *The Organizational Development Institute, Cleveland, OH*, 57–64.

Kilpatrick, Marcus, Hebert, Edward, & Bartholomew, John. 2005. College students' motivation for physical activity: Differentiating men's and women's motives for sport participation and exercise. *Journal of American College Health*, **54**(2), 87–94.

King, Dominic, Greaves, Felix, Exeter, Christopher, & Darzi, Ara. 2013. *'Gamification': Influencing health behaviours with games.*

Klenk, Saskia, Reifegerste, Doreen, & Renatus, Rebecca. 2017. Gender differences in gratifications from fitness app use and implications for health interventions. *Mobile Media & Communication*, **5**(2), 178–193.

Knell, Markus. 1999. Social comparisons, inequality, and growth. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 664–695.

Knell, Markus, & Stix, Helmut. 2020. Perceptions of inequality. *European Journal of Political Economy*, **65**, 101927.

Knez, Marc J, & Camerer, Colin F. 1995. Outside options and social comparison in three-player ultimatum game experiments. *Games and Economic Behavior*, **10**(1), 65–94.

Knight, Jonathan. 2002. Sexual stereotypes. *Nature*, **415**(6869), 254–256.

Kraus, Michael W, Rucker, Julian M, & Richeson, Jennifer A. 2017. Americans misperceive racial economic equality. *Proceedings of the National Academy of Sciences*, **114**(39), 10324–10331.

Krawczyk, Michał. 2010. A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution. *Journal of Public Economics*, **94**(1-2), 131–141.

Kuziemko, Ilyana, Buell, Ryan W, Reich, Taly, & Norton, Michael I. 2014. "Last-place aversion": Evidence and redistributive implications. *The Quarterly Journal of Economics*, **129**(1), 105–149.

Kuznets, Simon. 1955. Economic Growth and Income Inequality. *The American Economic Review*, **45**(1), 1–28.

Laenen, Tijs, & Roosma, Femke. 2022. Who should get what and why? Insights from welfare deservingness theory. *Pages 65–75 of: Solidarity and social justice in contemporary societies: An interdisciplinary approach to understanding inequalities*. Springer.

Lauderdale, Michael E, Yli-Piipari, Sami, Irwin, Carol C, & Layne, Todd E. 2015. Gender differences regarding motivation for physical activity among college students: A self-determination approach. *The Physical Educator*, **72**(5).

Layte, Richard, & Whelan, Christopher T. 2014. Who feels inferior? A test of the status anxiety hypothesis of social inequalities in health. *European Sociological Review*, **30**(4), 525–535.

Lester, Richard T, Ritvo, Paul, Mills, Edward J, Kariri, Antony, Karanja, Sarah, Chung, Michael H, Jack, William, Habyarimana, James, Sadatsafavi, Mohsen, Najafzadeh, Mehdi, *et al.* 2010. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): A randomised trial. *The Lancet*, **376**(9755), 1838–1845.

Levitt, Steven D, List, John A, Neckermann, Susanne, & Sadoff, Sally. 2016. The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, **8**(4), 183–219.

Leyens, Jacques-Philippe, Désert, Michel, Croizet, Jean-Claude, & Darcis, Catherine. 2000. Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin*, **26**(10), 1189–1199.

Liebkind, Karmela, Larja, Lisa, & Brylka, Asteria Anna. 2016. Ethnic and gender discrimination in recruitment: Experimental evidence from Finland. *Journal of Social and Political Psychology*.

Ludwig, Sandra, Fellner-Röhling, Gerlinde, & Thoma, Carmen. 2017. Do women have more shame than men? An experiment on self-assessment and the shame of overestimating oneself. *European Economic Review*, **92**, 31–46.

Mago, Shakun D, & Razzolini, Laura. 2019. Best-of-five contest: An experiment on gender differences. *Journal of Economic Behavior & Organization*, **162**, 164–187.

Mata, Jutta, Silva, Marlene N, Vieira, Paulo N, Carraça, Eliana V, Andrade, Ana M, Coutinho, Sílvia R, Sardinha, Luis B, & Teixeira, Pedro J. 2009. Motivational "spillover" during weight control: Increased self-determination and exercise intrinsic motivation predict eating self-regulation. *Health Psychology*, **28**(6), 709.

Matud, M Pilar. 2004. Gender differences in stress and coping styles. *Personality and Individual Differences*, **37**(7), 1401–1415.

Matzer, Franziska, Nagele, Eva, Lerch, Nikolaus, Vajda, Christian, & Fazekas, Christian. 2018. Combining walking and relaxation for stress reduction—A randomized cross-over trial in healthy adults. *Stress and Health*, **34**(2), 266–277.

McGinnity, Frances, Nelson, Jacqueline, Lunn, Pete, & Quinn, Emma. 2009. Discrimination in Recruitment: Evidence from a field experiment.

Meltzer, Allan H, & Richard, Scott F. 1981. A rational theory of the size of government. *Journal of Political Economy*, **89**(5), 914–927.

Mitchell, Terence R, Thompson, Leigh, Peterson, Erika, & Cronk, Randy. 1997. Temporal adjustments in the evaluation of events: The "rosy view". *Journal of Experimental Social Psychology*, **33**(4), 421–448.

Möbius, Markus M, Niederle, Muriel, Niehaus, Paul, & Rosenblat, Tanya S. 2022. Managing self-confidence: Theory and experimental evidence. *Management Science*, **68**(11), 7793–7817.

Molanorouzi, Keyvan, Khoo, Selina, & Morris, Tony. 2015. Motives for adult participation in physical activity: Type of activity, age, and gender. *BMC Public Health*, **15**, 1–12.

Mollerstrom, Johanna, Reme, Bjørn-Atle, & Sørensen, Erik Ø. 2015. Luck, choice and responsibility—An experimental study of fairness views. *Journal of Public Economics*, **131**, 33–40.

Monteil, Claudine. 2011. *Simone de Beauvoir et les femmes aujourd'hui.* Odile Jacob.

Morais Maceira, Helena. 2017. Economic benefits of gender equality in the EU. *Intereconomics*, **52**(3), 178–183.

Moss-Racusin, Corinne A, & Rudman, Laurie A. 2010. Disruptions in women's self-promotion: The backlash avoidance model. *Psychology of Women Quarterly*, **34**(2), 186–202.

Müller, Daniel, & Renes, Sander. 2021. Fairness views and political preferences: Evidence from a large and heterogeneous sample. *Social Choice and Welfare*, **56**(4), 679–711.

Murphy, Ryan O, Ackermann, Kurt A, & Handgraaf, Michel JJ. 2011. Measuring social value orientation. *Judgment and Decision making*, **6**(8), 771–781.

Nicholls, Nicky. 2022. Race, gender and willingness to compete. *Economics Letters*, **213**, 110382.

Niederle, Muriel, & Vesterlund, Lise. 2007. Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, **122**(3), 1067–1101.

Niehues, Judith. 2014. Subjective perceptions of inequality and redistributive preferences: An international comparison. *Cologne Institute for Economic Research. IW-TRENDS Discussion Paper*, **2**(1), 23.

Norton, Michael I, & Ariely, Dan. 2011. Building a better America—One wealth quintile at a time. *Perspectives on Psychological Science*, **6**(1), 9–12.

Nosek, Brian A, Banaji, Mahzarin R, & Greenwald, Anthony G. 2002. Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology*, **83**(1), 44–59.

Nyatsanza, Farai, McSorley, John, Murphy, Siobhan, & Brook, Gary. 2016. 'It's all in the message': the utility of personalised short message service (SMS) texts to remind patients at higher risk of STIs and HIV to reattend for testing—a repeat before and after study. *Sexually Transmitted Infections*, **92**(5), 393–395.

O'Connell, Stephen D, & Lang, Guido. 2018. Can personalized nudges improve learning in hybrid classes? Experimental evidence from an introductory undergraduate course. *Journal of Research on Technology in Education*, **50**(2), 105–119.

O'Donoghue, Ted, & Rabin, Matthew. 1999. Doing It Now or Later. *American Economic Review*, **89**(1), 103–124.

OECD. 2023. *Gender wage gap*.

Osberg, Lars, & Smeeding, Timothy. 2006. "Fair" inequality? Attitudes toward pay differentials: The United States in comparative perspective. *American Sociological Review*, **71**(3), 450–473.

Patel, Mitesh S, Volpp, Kevin G, Rosin, Roy, Bellamy, Scarlett L, Small, Dylan S, Fletcher, Michele A, Osman-Koss, Rosemary, Brady, Jennifer L, Haff, Nancy, Lee, Samantha M, *et al.* 2016. A randomized trial of social comparison feedback and financial incentives to increase physical activity. *American Journal of Health Promotion*, **30**(6), 416–424.

Peer, Eyal, Egelman, Serge, Harbach, Marian, Malkin, Nathan, Mathur, Arunesh, & Frik, Alisa. 2020. Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior*, **109**, 106347.

Petit, Emmanuel. 2022. Théorie des émotions et analyse économique: une revue 1. *Revue d'économie politique*, **32**(2), 181–215.

Phelps, Edmund S. 1972. The statistical theory of racism and sexism. *The American Economic Review*, **62**(4), 659–661.

Piketty, Thomas. 2021. *Une brève histoire de l'égalité*. Éditions du Seuil Paris.

Price, Joseph. 2008. Gender differences in the response to competition. *Industrial and Labor Relations Review*, **61**(3), 320–333.

Price, Joshua. 2010. The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, **29**(6), 901–910.

Putterman, Louis. 1997. Why have the rabble not redistributed the wealth? On the stability of democracy and unequal property. *Pages 359–393 of: Property Relations, Incentives*

and Welfare: Proceedings of a Conference held in Barcelona, Spain, by the International Economic Association. Springer.

Rawls, John. 1971. *A Theory of Justice*. Belknap Press.

Roberts, Benjamin J, Struwig, Jarè, Edlund, Jonas, & Lindh, Arvid. 2023. The ISSP 2019 social inequality module: country-comparative individual-level data on public beliefs about inequality and socioeconomic conditions over three decades. *International Journal of Sociology*, **53**(5), 333–350.

Rodgers, Anthony, Corbett, Tim, Bramley, Dale, Riddell, Tania, Wills, Mary, Lin, Ruey-Bin, & Jones, Mark. 2005. Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tobacco Control*, **14**(4), 255–261.

Rondina, Renante, Hong, Michael, Sarma, Sisira, & Mitchell, Marc. 2021. Is it worth it? Cost-effectiveness analysis of a commercial physical activity app. *BMC Public Health*, **21**, 1–10.

Royer, Heather, Stehr, Mark, & Sydnor, Justin. 2015. Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, **7**(3), 51–84.

Rozin, Paul, & Royzman, Edward B. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, **5**(4), 296–320.

Rudman, Laurie A, Moss-Racusin, Corinne A, Phelan, Julie E, & Nauts, Sanne. 2012. Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, **48**(1), 165–179.

Russell, Hayley C, Potts, Charlie, & Nelson, Emma. 2023. "If It's not on Strava it Didn't Happen": Perceived Psychosocial Implications of Strava use in Collegiate Club Runners. *Recreational Sports Journal*, **47**(1), 15–25.

Samuelson, William, & Zeckhauser, Richard. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty*, **1**, 7–59.

Sanfey, Alan G. 2009. Expectations and social decision-making: Biasing effects of prior knowledge on Ultimatum responses. *Mind & Society*, **8**, 93–107.

Schmidt, Ulrich, Friedl, Andreas, Eichenseer, Michael, & de Miranda, Katharina Lima. 2021. Social comparison and gender differences in financial risk taking. *Journal of Economic Behavior & Organization*, **192**, 58–72.

Sedikides, Constantine, & Green, Jeffrey D. 2009. Memory as a self-protective mechanism. *Social and Personality Psychology Compass*, **3**(6), 1055–1068.

Serres, Michel. 2017. *C'était mieux avant!* Le Pommier.

Sharon-David, Hilla, & Tenenbaum, Gershon. 2017. The effectiveness of exercise interventions on coping with stress: Research synthesis. *Studies in Sport Humanities*, **22**, 19–29.

Sittenthaler, Hanna M, & Mohnen, Alwine. 2020. Cash, non-cash, or mix? Gender matters! The impact of monetary, non-monetary, and mixed incentives on performance. *Journal of Business Economics*, **90**(8), 1253–1284.

Slaunwhite, Jason M, Smith, Steven M, Fleming, Mark T, & Fabrigar, Leandre R. 2009. Using normative messages to increase healthy behaviours. *International Journal of Workplace Health Management*, **2**(3), 231–244.

Sommet, Nicolas, Elliot, Andrew J, Jamieson, Jeremy P, & Butera, Fabrizio. 2019. Income inequality, perceived competitiveness, and approach-avoidance motivation. *Journal of Personality*, **87**(4), 767–784.

Sorensen, Elaine. 1990. The crowding hypothesis and comparable worth. *Journal of Human Resources*, 55–89.

Spoor, Jennifer R, & Schmitt, Michael T. 2011. "Things are getting better" isn't always better: Considering women's progress affects perceptions of and reactions to contemporary gender inequality. *Basic and Applied Social Psychology*, **33**(1), 24–36.

Steiner, John F, Shainline, Michael R, Dahlgren, Jennifer Z, Kroll, Alan, & Xu, Stan. 2018. Optimizing number and timing of appointment reminders: A randomized trial. *Am J Manag Care*, **24**(8), 377–384.

Stiglitz, Joseph. 2015. *The great divide*. Penguin Uk.

Strain, Tessa, Flaxman, Seth, Guthold, Regina, Semenova, Elizaveta, Cowan, Melanie, Riley, Leanne M, Bull, Fiona C, & Stevens, Gretchen A. 2024. National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: A pooled analysis of 507 population-based surveys with 5· 7 million participants. *The Lancet Global Health*, **12**(8), e1232–e1243.

Sutter, Matthias, & Glätzle-Rützler, Daniela. 2015. Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, **61**(10), 2339–2354.

Tammi, Timo. 2013. Dictator game giving and norms of redistribution: Does giving in the dictator game parallel with the supporting of income redistribution in the field? *The Journal of Socio-Economics*, **43**, 44–48.

Trautmann, Sebastian, Rehm, Jürgen, & Wittchen, Hans-Ulrich. 2016. The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO reports*, **17**(9), 1245–1249.

Trivers, Robert L. 1972. *Parental investment and sexual selection*. Routledge.

Tu, Rungting, Hsieh, Peishan, & Feng, Wenting. 2019. Walking for fun or for "likes"? The impacts of different gamification orientations of fitness apps on consumers' physical activities. *Sport Management Review*, **22**(5), 682–693.

Tufano, James T, & Karras, Bryant T. 2005. Mobile eHealth interventions for obesity: A timely opportunity to leverage convergence trends. *Journal of Medical Internet Research*, **7**(5), e467.

Ugur, Zeynep B. 2021. How does inequality hamper subjective well-being? The role of fairness. *Social Indicators Research*, **158**(2), 377–407.

UN Women. 2023. *Progress on the Sustainable Development Goals: The gender snapshot 2023*. Tech. rept.

Uslaner, Eric M, & Brown, Mitchell. 2005. Inequality, trust, and civic engagement. *American Politics Research*, **33**(6), 868–894.

Van Veldhuizen, Roel. 2022a. Gender differences in tournament choices: Risk preferences, overconfidence, or competitiveness? *Journal of the European Economic Association*, **20**(4), 1595–1618.

Van Veldhuizen, Roel. 2022b. Gender Differences in Tournament Choices: Risk Preferences, Overconfidence, or Competitiveness? *Journal of the European Economic Association*, **20**(4), 1595–1618.

Vervloet, Marcia, Linn, Annemiek J, van Weert, Julia CM, De Bakker, Dinny H, Bouvy, Marcel L, & Van Dijk, Liset. 2012. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. *Journal of the American Medical Informatics Association*, **19**(5), 696–704.

Vlaev, Ivo, King, Dominic, Darzi, Ara, & Dolan, Paul. 2019. Changing health behaviors using financial incentives: A review from behavioral economics. *BMC Public Health*, **19**(1), 1–9.

Waldfogel, Hannah B, Sheehy-Skeffington, Jennifer, Hauser, Oliver P, Ho, Arnold K, & Kteily, Nour S. 2021. Ideology selectively shapes attention to inequality. *Proceedings of the National Academy of Sciences*, **118**(14), e2023985118.

Walsh, Jane C, Corbett, Teresa, Hogan, Michael, Duggan, Jim, & McNamara, Abra. 2016. An mHealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. *JMIR mHealth and uHealth*, **4**(3), e5227.

Wilkinson, Richard G, & Pickett, Kate E. 2017. The enemy between us: The psychological and social costs of inequality. *European Journal of Social Psychology*, **47**(1), 11–24.

Willis, Guillermo B, García-Sánchez, Efraín, Sánchez-Rodríguez, Ángel, García-Castro, Juan Diego, & Rodríguez-Bailón, Rosa. 2022. The psychosocial effects of economic inequality depend on its perception. *Nature Reviews Psychology*, **1**(5), 301–309.

Wood, Joanne V. 1996. What is social comparison and how should we study it? *Personality and Social Psychology Bulletin*, **22**(5), 520–537.

Wood, Joanne V, & Taylor, Kathryn L. 1991. Serving self-relevant goals through social comparison. *Pages 23–49 of: Social comparison*. Routledge.

Wood, Martin, Hales, Jon, Purdon, Susan, Sejersen, Tanja, & Hayllar, Oliver. 2009. A test for racial discrimination in recruitment practice in British cities. *Department for Work and Pensions Research Report*, **607**.

World Health Organization. 2019. *Global action plan on physical activity 2018-2030: more active people for a healthier world*. World Health Organization.

Wozniak, David, Harbaugh, William T, & Mayr, Ulrich. 2009. *Choices about competition: Differences by gender and hormonal fluctuations, and the role of relative performance feedback*. Tech. rept. *Available at SSRN 1564895*.

Wozniak, David, Harbaugh, William T, & Mayr, Ulrich. 2014. The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*, **32**(1), 161–198.

Xu, Xiaogeng, Metsälampi, Satu, Kirchler, Michael, Kotakorpi, Kaisa, Matthews, Peter Hans,

& Miettinen, Topi. 2023. *Which income comparisons matter to people, and how? Evidence from a large field experiment.* Tech. rept. Working Papers in Economics and Statistics.

Yeung, Catherine, Ho, Teck-Hua, Sato, Ryoko, Lim, Noah, Van Dam, Rob M, Tan, Hong-Chang, Tham, Kwang-Wei, & Ali, Rehan. 2024. Cash incentives for weight loss work only for males. *Behavioural Public Policy*, **8**(2), 279–299.

Yin, Qiuju, Li, Lun, Yan, Zhijun, & Guo, Chenxi. 2022. Understanding the effects of self-peer-platform incentives on users' physical activity in mobile fitness apps: The role of gender. *Information Technology & People*, **35**(3), 1054–1072.

Liste des figures

1 Répartition des revenus en France 9

2 "À quel type de société correspond la France ?" 11

1.1 Timeline of the experiment (Part I) 28

1.2 Example of the performance gap's scale 31

1.3 Percentage of subjects choosing the competition in Stage 3, by gender and condition 38

1.4 Distribution of the switching point, by outcome in Stage 3T and gender of the subject. 42

1.5 Average switching point, by outcome in Stage 3T and condition 43

1.6 Performance in Stage 3T and Stage 4aT, by outcome in Stage 3T and condition 47

1.B1 Average performance in each stage of the game, by gender and conditional on the decision to compete made at Stage 3 52

1.C1 Distribution of the WTP at Stage 4b, by gender 53

1.C2 Average performance in Stage 3PR and Stage 4bT, by condition 54

2.1 Percentage of Option B, by treatment 96

2.2 Percentage of Player 2s who perceive the allocation received (Option A) as *fair*, by treatment 100

2.A1 Distribution of answers to ISSP's survey 105

2.A2 Distribution of choice among prosocial subjects, by treatment 106

2.C1 Distribution of the categories, by treatment, according to the research assistant 114

3.1 Timeline of the project 130

3.2 Share of our users who open *WeWard*, by day of the week. 138
3.3 Time trend of share of subjects who open *WeWard* 139
3.4 Average number of steps per week by treatment group. 148
3.5 Average number of steps, by day of the week. 150
3.6 *Daily Steps*, by treatment group. 150
3.7 *Daily Steps*, by category. 153
3.8 *Weekly Steps*, by category. 153
3.A1 Screenshot of the app 157
3.A2 Fraction of users in the population, by French region 158
3.A3 Proportion of our users' sample who open *WeWard*, by gender and the day of the week 159
3.A4 *Weekly Steps* across the experiment window, by the category of walkers. . . . 159

4.1 Share of users who open *WeWard*, by day of the week 177
4.2 Probability of validating steps in relation to the thresholds, by gender 178
4.3 RDD plots for all thresholds 179
4.A1 Screenshot of the app 182
4.A2 Fraction of users in the population, by French region 183
4.B1 Gender gap in the distribution of the number of steps 186
4.B2 Gender gap in the number of recorded steps by day of the week 186
4.B3 Evolution of the weekly steps across the observation weeks, by gender 187
4.C1 Time trend of share of subjects who open *WeWard*, by gender 190
4.C2 RDD plot for the **first threshold** 190
4.C3 RDD plot for the **second threshold** 191
4.C4 RDD plot for the **third threshold** 191
4.C5 RDD plot for the **fourth threshold** 192
4.C6 RDD plot for the **fifth threshold** 192
4.C7 RDD plot for the **sixth threshold** 193

4.C8 Confidence intervals for the estimated coefficients of Table 4.1, by gender . . 193

Liste des tableaux

1.1 Summary of descriptive statistics 37

1.2 Determinants of entry into competition in Stage 3. 40

1.3 Determinants of Switching point 45

1.4 Determinants of performance in Stage 4aT 48

1.A1 Answers to the question: "How do you perceive these names?" (N=112) . . . 52

1.C1 Willingness to pay to avoid competition in Stage 4b 55

1.D1 Answers to the question: "What is the presumed topic of our study?" 57

1.D2 Answers to the question: "Why did you (not) choose the tournament at Stage 3?" 57

1.E1 Switching point (including subjects who switch more than once) 58

1.E2 Switching point with with confidence 59

1.E3 Switching point (only subjects who correctly guess the gender of their opponent) 60

1.E4 Switching point (OLS estimation) 61

2.1 Options of distributions 88

2.2 Summary of descriptive statistics 94

2.3 The impact of the treatments on the choice of allocation 98

2.B1 Determinants of choice of Option B in each treatment 108

2.B2 The impact of the treatments on the fairness perceptions 110

2.C1 Answers to the question "Why do you think the decision is (un)fair?" 113

3.1 Summary statistics 131

3.2 Effects on the probability of opening the app: **Treatment** vs. **Pre-treatment** 140

3.3 Effects on the probability of opening the app: **Post-treatment** vs. **Treatment** . . . 141

3.4 Effects on *Number of connections*: **Treatment** vs. **Pre-treatment** 144

3.5 Effects on *Number of connections*: **Post-treatment** vs. **Treatment** 145

3.6 Effects on *Steps walked*: **Treatment** vs. **Pre-treatment** 147

3.7 Effects on *Weekly Steps*: **Treatment** vs. **Pre-treatment** 149

3.8 Effects on *Daily Steps*: **Treatment** vs. **Pre-treatment** 151

3.9 Effects on *Weekly steps* by initial level of walking: **Treatment** vs. **Pre-treatment** . . . 155

3.B1 Summary statistics 160

3.B2 Summary statistics: Strongly committed 160

3.B3 Effects on the probability of opening the app: **Post-treatment** vs. **Pre-treatment** . . . 161

3.B4 Effects on *Number of connections*: **Post-treatment** vs. **Pre-treatment** . . 162

3.B5 Effects on *Steps walked*: **Post-treatment** vs. **Treatment** 163

3.B6 Effects on *Steps walked*: **Post-treatment** vs. **Pre-treatment** 164

3.B7 Effects on *Weekly Steps*: **Post-treatment** vs. **Treatment** 165

3.B8 Effects on *Weekly Steps*: **Post-treatment** vs. **Pre-treatment** 166

3.B9 Effects on *Daily Steps*: **Post-Treatment** vs. **Treatment** 167

3.B10 Effects on *Daily Steps*: **Post-Treatment** vs. **Pre-Treatment** 168

3.B11 Effects on *Weekly Steps* by initial level of walking: **Post-treatment** vs. **Treatment** . . . 169

3.B12 Effects on *Weekly Steps* by initial level of walking: **Post-treatment** vs. **Pre-treatment** . . . 170

4.1 The effects of thresholds on the probability of validating steps 180

4.B1 Summary statistics *We Ward* dataset 185

4.B2 Baromètre de la pratique sportive - Édition 2020 188

4.B3 Enquête nationale sur les pratiques sportives - Édition 2020 189

4.C1 The effects of thresholds on the probability of validating steps with a cutoff to **200 steps** 194

4.C2 The effects of thresholds on the probability of validating steps with a cutoff to **700 steps** 195

4.C3 The effects of being above the threshold for males 196

4.C4 The effects of being above the threshold for females 196