Essays on the Behavioral Economics of Motivated Memory

Charlotte Saucet

Numéro National de Thèse: **2019LYSEN069**

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée par
l'École Normale Supérieure de Lyon

École Doctorale n^o486
Sciences Économiques et de Gestion

Spécialité de Doctorat : Sciences Économiques

Soutenue publiquement le 10 décembre 2019, par:

Charlotte SAUCET

Essays on the Behavioral Economics of Motivated Memory

Essais en économie comportementale sur la mémoire motivée

Devant le jury composé de:

Paul Seabright	-	Professeur, Toulouse School of Economics, *Rapporteur*
Sigrid Suetens	-	Professeure, Tilburg School of Economics and Management, *Rapporteure*
Benoît Tarroux	-	Professeur, Université de Lyon, *Examinateur*
Florian Zimmermann	-	Professeur assistant, Briq et Université de Bonn, *Examinateur*
Marie Claire Villeval	-	Directrice de recherche CNRS, Université de Lyon, *Directrice de thèse*

Ecole Normale Supérieure de Lyon is not going to give any approbation or disapprobation about the thoughts expressed in this dissertation. They are only the author's ones and need to be considered such as.

Remerciements

Ces quatre années de doctorat constituent pour moi, au-delà de l'apprentissage de la recherche et du métier de chercheur, un vrai voyage initiatique. L'initiation était plurielle, stimulante, intense. Éprouvante parfois, formatrice toujours. Chercher, c'est accepter l'incertitude du résultat, avec la frustration, le doute, et la remise en question que cela peut parfois engendrer. Chercher, c'est apprendre la patience, l'endurance, et l'humilité. C'est lire, beaucoup, écrire, toujours, compter aussi, et savoir bien le faire. Mais chercher, c'est avant tout explorer des horizons inconnus, rencontrer des personnes passionnées et passionnantes, mélanger les disciplines, se questionner, être curieux, apprendre, découvrir, et tenter de comprendre. Jamais mon voyage n'aurait été si formateur sans le soutien d'un grand nombre de personnes que je tiens ici à remercier chaleureusement.

Merci à vous Marie Claire. Je pourrais m'épancher longuement sur vos immenses qualités de chercheure. Mais, au-delà du profond respect que je porte à celle que vous êtes, passionnée et exigeante, c'est aussi la personne que je voudrais honorer. Merci pour votre oreille toujours attentive, votre considération et votre bienveillance. Merci de m'avoir accordé maintes fois ces heures de discussions, enrichissantes et humaines, dans votre emploi du temps pourtant si surchargé. Merci de m'avoir ouvert tant de portes. Merci, surtout, de m'avoir fait confiance.

Merci aux membres de mon Jury de thèse, Benoît Tarroux et Florian Zimmermann, d'avoir accepté d'en faire partie. Merci à Sigrid Suetens et à Paul Seabright d'avoir accepté le rôle de rapporteur. C'est un honneur et un plaisir de présenter mon travail devant vous.

Merci à mes deux co-auteurs, Fabio Galeotti et Alberto Prati. Fabio, merci pour ton expertise, ta réactivité, et ton optimisme volontairement rassurant. C'est un luxe pour un doctorant de travailler avec un chercheur confirmé qui met autant les mains dans le cambouis. Alberto, ton envie, ta curiosité et ta sensibilité philosophe sont à la fois inspirantes et nourrissantes. Je suis ravie que nos intérêts communs pour l'étude de la mémoire en économie se soient rencontrés.

Merci à tous les chercheurs qui m'ont accueillie dans leur laboratoire pendant mes divers séjours à l'étranger. Merci à Gary Charness, à Zachari Grossman, à Ryan Oprea et à Peter Khun de l'Université de Californie Santa Barbara, aux États-Unis. Assister à vos cours, spécialisés, rigoureux et souvent très animés, est un de mes meilleurs souvenirs académiques. Gary, côtoyer un chercheur d'une telle trempe à l'aube de ma thèse a été plus que formateur, tant intellectuellement qu'humainement. Je n'oublierai pas la métaphore des étoiles. Merci à Luis Santos Pinto d'HEC Lausanne, à l'équipe du CREED à Amsterdam, notamment à Jeroen Van de Ven et Joël Van der Weele, et à celle de la TSE à Toulouse. Je tiens également à remercier Maud Frot et Jane Plailly du Centre de recherches en Neurosciences de Lyon (CNRL) et Gaën Plancher du Laboratoire d'Étude des Mécanismes Cognitifs (EMC) de l'Université Lyon 2, toutes trois chercheures confirmées, d'avoir fait confiance à la jeune doctorante que j'étais pour venir compléter l'équipe d'organisation d'une conférence internationale sur la mémoire. C'était une belle réussite.

Merci à l'ensemble du personnel administratif du laboratoire, pour votre sympathie et votre professionnalisme. Travailler dans de telles conditions, humaines et matérielles, est une chance. Merci à la directrice du Gate, Sonia Paty, d'y veiller au quotidien. Un merci particulier à Aude pour nos échanges de couloirs complices, et à Quentin, même si je me sentirai toujours un peu sourde à tes côtés.

Merci à tous les doctorants et post-doctorants du laboratoire, et plus particulièrement à Claire, Clément, Fortuna, Liza, Julien, Marius, Morgan, Rémi, Sorravich, Tatiana, Thomas, Tidiane, Valentin, Vincent et Yohann (par ordre alphabétique), qui ont été de si bons compagnons d'aventure. C'était fort agréable de faire partie de cette joyeuse bande de lurons toujours très motivée lorsqu'il s'agit de lever son

verre, mais aussi et surtout de s'entraider. Avis à tous : mon titre de championne de ping-pong du Gate est remis en jeu !

Merci à ma famille, qui a essayé, puis accepté de ne pas forcément comprendre. Merci d'avoir su rester si proche et si aimante malgré la distance que cela a parfois pu instaurer, et merci pour l'inconditionnalité de votre amour : il n'y a rien de plus structurant que cela.

Merci à mes amis. Rien de tel que les moments passés à vos côtés à partager, à rire, à refaire le monde et nous refaire nous-mêmes, pour prendre du recul et revenir à l'essentiel. Avoir des amis comme vous est une ressource inestimable.

Enfin, merci à toi F. Toi qui as partagé au quotidien mes joies, mes doutes, et l'immensité de mon spectre émotionel, sans n'en être jamais effrayé. Ton soutien indéfectible est une force. Ton amour une richesse. De nouvelles aventures nous attendent...

À tous, merci. Grandir à vos côtés est une chance inouïe.

Acknowledgements

These past four years of my doctoral studies have been more than simply an education in research and academia. For me, they were an intense, stimulating, sometimes grueling –but always formative– journey. A journey that taught me to deal with uncertainty and to wrestle with my doubts. It instilled patience, endurance, and humility. Above all, it was an ongoing exploration, one where you meet passionate and thereby fascinating people, mix disciplines, question yourself, and unleash your curiosity. It would not have been such a formative experience without the support of many people whom I would like to thank here.

Without any hesitation, I want to first give thanks to Marie Claire. Marie Claire, I could spend days talking about your immense skill as a researcher, entirely devoted, rigorous, and demanding. But beyond the researcher, it is the person that I would like to honor. Thank you for your attentive listening, your consideration, and your kindness. Thank you for the many hours we shared engaged in enriching and refined discussions. Thank you for opening so many doors for me. Above all, thank you for your unwavering faith in me whenever I doubted myself. I wish all PhD students could have a supervisor with that deep understanding of human relations.

Thank you to Sigrid Suetens and Paul Seabright, who have kindly accepted the role of discussants for this thesis, and Benoît Tarroux and Florian Zimmermann, who agreed to be members of the jury. It is an honor and pleasure to present my work to you.

Thank you to my two co-authors, Fabio Galeotti and Alberto Prati. Fabio, I appreciated your expertise, your promptness, and your reassuring optimism. It

is a luxury for a PhD student to work with a researcher who never hesitates on getting his hands dirty. Alberto, your curiosity and philosophical nuance are both inspiring and nourishing. I am glad we met and bonded over our common interest in memory.

Thank you to all the researchers who welcomed me to their laboratory during my visits abroad. Thanks to Gary Charness, Zachari Grossman, Ryan Oprea and Peter Khun from the University of California, Santa Barbara, in the United States. My memories of attending your specialized, rigorous, and interactive courses are some of the best I have of academia. Gary, I grew both as an intellectual and as a human from being around a researcher of your disposition at the dawn of my thesis. I will never forget your "stars" metaphor. Thanks to Luis Santos Pinto from HEC Lausanne, to the CREED team in Amsterdam –especially Joël Van der Weele and Jeroen Van de Ven with whom talking about research is above all thinking "out of the box"–, and to the TSE team in Toulouse.

I would also like to thank Maud Frot and Jane Plailly of the Lyon Neurosciences Research Center (CNRL) and Gaën Plancher of the Laboratory for the Study of Cognitive Mechanisms (EMC) of the University Lyon 2, all three confirmed researchers, for putting their trust in the young doctoral student I was, and inviting me to join their organizing team for an international conference on memory. It was a great success.

Thank you to all the administrative staff at Gate for your sympathy and professionalism. I appreciate how lucky I have been to work in the material and humane conditions present there. Thank you to the Gate's director, Sonia Paty, for ensuring they continue every single day. A special thanks to Aude for our involvement; and to Quentin, even if I will always feel a little deaf at your side.

Thanks to all the doctoral and post-doctoral students in the laboratory, and more particularly to –in alphabetical order– Claire, Clément, Fortuna, Liza, Julien, Marius, Morgan, Rémi, Sorravich, Tatiana, Thomas, Tidiane, Valentin, Vincent and Yohann, who were such good companions on this adventure. I enjoyed being part of a group of people who are always very motivated not just when it comes to

having a toast, but also when we need each other's help. Beware: my title of Gate's Ping-Pong Champion is now back in play!

Thank you to my family, who tried to understand and finally accepted that they could not. Thank you for staying so close and loving despite the distance it has sometimes created, and thank you for your unconditional love. There is nothing more foundational and bracing than that.

Thanks to my friends. There is nothing better on earth to step back and get back to what is essential than the moments spent at your side talking, laughing, playing, remaking the world and exploring our complexity as human beings. Having friends like you is an invaluable resource.

Finally, thank you F. You, who shared my joys, my doubts, and the immensity of my emotional spectrum on a daily basis, without ever being afraid of them. Your unwavering support is a gift. Your love a treasure. New adventures await us...

Avant-propos

La mémoire parle à tout le monde. On la perd, on la recouvre, on la muscle. Bien que l'utilisation d'un terme unique puisse sembler suggérer que la mémoire fonctionne comme un système unitaire, les recherches scientifiques sur la mémoire, qui ont commencé il y a environ cent ans, ont montré que la mémoire n'est pas un, mais plusieurs systèmes (Baddeley, 1997). Les systèmes varient en nature (mémoire épisodique, mémoire procédurale, mémoire sémantique, etc.), en codage (mémoire visuelle, mémoire olfactive, mémoire sensorielle), en durée et en capacité de stockage (de quelques secondes à la mémoire à long terme) ; et en échelle (puisque la mémoire n'est pas seulement individuelle mais peut être collective).

La mémoire est dynamique, flexible, multidimensionnelle et donc complexe. C'est pourquoi la plupart des spécialistes parlent de *mémoires* plutôt que de *la mémoire*. Si certains d'entre eux, en dehors de la sphère économique, venaient à lire les études présentées ici, comme probablement toute autre étude économique sur la mémoire, ils se moqueraient probablement (s'ils ne s'outragent pas) de la façon dont nous considérons la mémoire comme une boîte noire. Dans notre cadre, les individus reçoivent de l'information (de leurs propres actions ou d'un feedback externe) et doivent ensuite s'en souvenir. Nous reconnaissons volontiers qu'un tel cadre est restrictif à bien des égards. Notamment, outre le fait d'ignorer les mécanismes cérébraux (en particulier les réseaux neuronaux et la plasticité cellulaire

permettant les différents processus de consolidation et de re-consolidation de la mémoire dans le temps, ce qui est d'un grand intérêt pour les neuroscientifiques), il se concentre sur un système très spécifique de la mémoire –la mémoire épisodique–, parmi bien d'autres qui ont été identifiés (voir Figure 1). L'utilisation du terme *mémoire* en économie et tout au long de cette thèse fait toujours référence au système épisodique de la mémoire. Ce dernier correspond à la mémoire d'événements autobiographiques, comme le souvenir de votre premier jour d'école ou le jour où vous avez reçu un rejet ferme d'un journal de rang un. Elle est différente de la mémoire sémantique qui renvoie à des connaissances plus générales que nous avons accumulées tout au long de notre vie, comme la connaissance de la capitale de la France ou le classement des cinq premières revues en économie. Ces systèmes appartiennent à la mémoire déclarative, qui se réfère au souvenir conscient et intentionnel d'événements et de faits passés (Ullman, 2004). D'autres systèmes de mémoire s'appuient sur la mémoire implicite qui, en revanche, est acquise et utilisée inconsciemment (Schacter, 1987).

Cette vision restrictive de la mémoire en économie étant reconnue, nous sommes convaincus que l'introduction de la mémoire dans cette discipline demeure d'une importance capitale. L'économiste n'a pas nécessairement besoin de savoir ce qu'il y a dans la boîte noire de la mémoire. Ce qui importe, en revanche, c'est de savoir comment et dans quelle mesure la mémoire intervient dans le processus de prise de décision. La mémoire des expériences passées est l'une des principales sources d'information sur nous-mêmes et sur le monde qui nous entoure. Explorer comment les individus récupèrent ces informations permet de mieux comprendre comment ils forment et actualisent leurs croyances et apprennent de leurs expériences pour prendre des décisions. L'étude d'une mémoire (sélective) de l'information permet également de mieux appréhender l'émergence de biais de perception et de

comportement, comme la sur-confiance en soi, qui peuvent avoir des implications majeures sur la qualité des choix. En outre, dès lors que "tout ce qui fausse la capacité des individus à se souvenir d'un événement passé fausse leur évaluation des probabilités futures" (Hammond et al., 2006), explorer les déterminants comportementaux des biais de mémoire peut permettre de mieux comprendre comment les individus déterminent leurs attentes et forment leurs anticipations. L'introduction de la mémoire dans le processus de prise de décision des individus met en évidence un candidat nouveau et sérieux pour explorer la dynamique des croyances motivées. Elle constitue une piste nouvelle en économie permettant d'expliquer l'existence de décisions sous-optimales. Ainsi, tout comme les philosophes, psychologues, sociologues ou neuroscientifiques explorent les concepts d'identité, d'éthique ou de communication de manières très différentes, les économistes ne doivent pas avoir peur d'explorer la mémoire humaine de manière restrictive ou simplifiée dès lors que cela leur permet de mieux comprendre comment les individus prennent des décisions dans un monde aux ressources limitées, tant matérielles que cognitives.

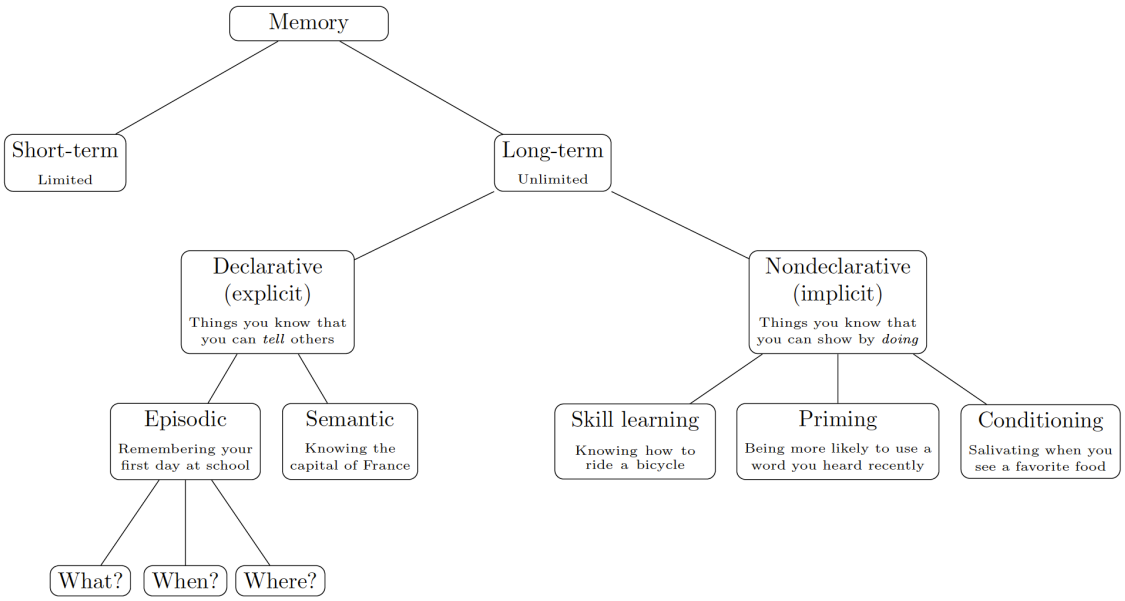


Figure 1: Les systèmes de la mémoire (adapté de Raslau et al. 2014)

Résumé de la thèse

Une des caractéristiques fondamentales de l'être humain est de construire des croyances, de préférence positives, sur lui-même et sur le monde qui l'entoure. Les individus chérissent des croyances qui leurs sont favorables ou désirables, et déploient de l'énergie et des efforts pour les protéger. À titre d'exemples, nous pouvons ainsi penser aux parents qui aiment à croire que leurs enfants sont particulièrement en avance sur leur développement, à un candidat se convaincant d'être plus compétent et adapté au poste que les autres candidats lors d'un entretien d'embauche, ou à une personne au régime se persuadant volontiers que ce n'est pas une petite glace à la crème qui fera la différence. Le besoin d'avoir des croyances positives sur soi et sur le monde qui nous entoure peut aussi s'illustrer à l'échelle macroéconomique. Un nombre significatif d'individus –dont certains dirigeants politiques– continuent de croire, en dépit des innombrables preuves scientifiques, que le réchauffement climatique n'est que partiellement dû à l'activité humaine et que les conséquences en découlant seraient minimes. Dans un autre domaine, la prédominance des religions dans la plus grande partie du globe atteste du besoin de croire des individus, sans que ces croyances ne soient nécessairement fondées sur des faits tangibles. À la lumière de ces exemples, de nombreuses études montrent que le plus souvent les croyances se forment et s'ajustent non pas façon neutre et objective, c'est à dire d'après un processus rationnel (bayésien) de traitement de l'information, mais en partie pour répondre à d'importants besoins psychologiques ou émotionnels. Les

individus recueillent et interprètent l'information de façon à confirmer ce qu'ils souhaitent croire. Dès lors, les croyances peuvent être considérées comme motivées (Bénabou, 2015).

Cette conception des croyances comme un "bien dans lequel les gens investissent" (Bénabou and Tirole, 2011) a profondément remis en question la vision de l'homo oeconomicus (Mill, 1874), censé former et ajuster ses croyances d'après un traitement rationnel de l'information. En réalité, tout comme Schelling (1987) décrit "l'esprit [humain] comme un organe consommateur", les individus semblent tirer de l'utilité non seulement de la consommation de biens ou d'expériences comme c'est le cas en économie standard, mais aussi de la consommation de croyances. Face à ce constat, les économistes comportementaux ont introduit le concept d' "utilité fondée sur les croyances" (belief-based utility) dans lequel les individus peuvent retirer de l'utilité de leurs croyances, même si ces dernières sont inexactes et nuisent à une prise de décision optimale.

Les mécanismes, cependant, sont complexes, et les individus ne croient pas *simplement* ce qu'ils ont envie de croire. Les croyances doivent être motivées de façon suffisamment plausible pour être tenues. Pour répondre à cette demande de croyances motivées, les individus doivent donc développer des stratégies leur permettant de se défendre de façon crédible face à des preuves ou des informations indésirables et, inversement, d'alimenter des croyances qui leurs sont favorables[1]. La question est : comment ? Les économistes se sont penchés sur cette question depuis des années, et deux types de stratégies ont principalement été explorées.

[1]On suppose ici que l'information qui est disponible aux individus n'est pas manipulée *per se*. Il s'agit d'une représentation restrictive de la réalité puisqu'il ne semble pas infondé de penser que des institutions telles que les gouvernements, les médias, les partis politiques et/ou les entreprises peuvent parfois avoir intérêt à fournir des informations ex-ante déformées ou fausses.

Ex ante, les individus préfèrent parfois le “bonheur de l’ignorance” au “pouvoir de la connaissance”. Ils évitent ou acquièrent l’information de façon sélective, même lorsque celle-ci est gratuite et permettrait d’améliorer la prise de décision (Golman et al., 2017). À titre d’exemples, Karlsson et al. (2009) ont montré que des investisseurs actualisent très fréquemment la valeur de leur portefeuille d’actions lorsque le marché est à la hausse, mais cessent de la regarder lorsque le marché est à la baisse. Dans le domaine médical, Oster et al. (2013) ont montré qu’une part significative des individus étudiés, susceptibles de contracter la maladie d’Huntington, refusent des tests de dépistages, pourtant gratuits, pour ne pas avoir à être confrontés à une éventuelle mauvaise nouvelle. Les stratégies d’évitement de l’information produisent de l’incertitude qui impacte les décisions des individus. De nombreuses études ont montré que les individus ignorent délibérément certaines informations indésirables pour s’autoriser à agir de façon intéressée. En utilisant un jeu du dictateur, Dana et al. (2007) ont montré que les dictateurs exploitent l’incertitude sur les conséquences de leurs décisions comme une marge de manoeuvre pour agir de façon plus égoïste que s’ils étaient informés (voir aussi Grossman, 2014; Kajackaite, 2015; Grossman and Van Der Weele, 2017). Les individus donnent également moins aux associations caritatives en présence de risque (Exley, 2015; Garcia et al., 2018) et d’incertitude (Garcia et al., 2018) que lorsque l’issue de leur décision est certaine. Fait intéressant, il n’y a pas qu’une demande d’ignorance d’information mais également une offre d’ignorance. Dans une expérience de laboratoire, Shalvi et al. (2019) ont montré que certains décideurs évitaient les conseillers qui leur transmettaient des informations indésirables (côté demande), mais également que la majorité des conseillers eux-mêmes supprimaient les informations indésirables lorsqu’ils transmettaient des informations aux décideurs (côté offre).

Ex post, la précision n’est pas non plus toujours l’objectif principal qui sous-tend la formation des croyances. Plusieurs études ont montré que les individus

mettent à jour leurs croyances de façon non-bayésienne (Rabin and Schrag, 1999). À titre d'exemple, Eil and Rao (2011) et Mobius et al. (2011) ont montré que les individus mettent à jour leurs croyances de façon bayésienne lorsqu'ils reçoivent de bonnes nouvelles concernant leur beauté et/ou leur intelligence, mais qu'ils font preuve de conservatisme dans l'ajustement de leurs croyances après avoir reçu de mauvaises nouvelles. En utilisant également un test de QI, Zimmermann (2019) a constaté que les croyances sont ajustées dans la bonne direction lorsqu'elles sont élicitées directement après avoir reçu un feedback, mais qu'elles sont mises à jour de façon asymétrique lorsqu'elles sont élicitées un mois après avoir reçu le feedback. Plus précisément, les croyances des participants qui ont reçu un feedback positif demeurent ajustées à la hausse, alors que les croyances des participants qui ont reçu un feedback négatif ont tendance à retourner à leur état initial. Outre cette distorsion de l'information, les individus peuvent alimenter leur besoin de croyances motivées en rejetant la faute sur les autres pour leurs actions (Bartling and Fischbacher, 2011; Oexl and Grossman, 2013), en se déresponsabilisant (Foerster and Van der Weele, 2018) ou en se trouvant des excuses consistant par exemple à minimiser les conséquences de leurs actes et/ou réinterpréter les circonstances de leur comportement (Bénabou et al., 2018; Foerster and van der Weele, 2018).

Bien que l'évitement et la mise à jour biaisée de l'information aient été des mécanismes largement étudiés en économie expérimentale, une dernière stratégie dont disposent les individus a longtemps été délaissée. Même lorsque l'information est reçue et mise à jour, les individus peuvent, en dernier recours, l'oublier. Les individus peuvent avoir tendance à oublier les informations indésirables et, inversement, fournir d'importants efforts lorsqu'il s'agit de se souvenir d'informations plaisantes ou valorisantes. En d'autres termes, le temps qui passe pourrait constituer une opportunité d'oublier ou de déformer ce dont les individus préfèrent ne pas se souvenir. Ce désintérêt pour la mémoire en économie semble très surprenant

lorsque l'on considère que la plupart des croyances qui sous-tendent nos décisions sont issues de notre mémoire (Tranel et al., 1994; Bechara et al., 1998). En fait, le souvenir que nous avons des événements passés est, sinon la première, l'une des principales sources d'information. La prise en compte de la mémoire dans le processus de prise de décision semble donc d'une importance cruciale. Elle met en lumière un candidat nouveau et sérieux permettant d'explorer la dynamique des croyances motivées.

Cette thèse cherche à déterminer si les individus manipulent leur mémoire pour soutenir leur désir de croyances motivées. Elle teste expérimentalement l'existence et la force de la mémoire motivée dans trois contextes économiquement pertinents que sont les préférences sociales, la performance individuelle et les décisions malhonnêtes ou immorales. Elle fournit un nouvel ensemble de preuves montrant que, même lorsque l'information a été reçue et mise à jour, une dernière stratégie à la disposition des individus consiste effectivement à oublier l'information qui menace leurs croyances.

> I have done this, says my memory. I cannot have done that, says my pride, remaining inexorable. Finally, memory yields.
>
> Nietzsche, 1886, *Beyond Good and Evil*

La mémoire a longtemps été considérée comme un espace de stockage dans lequel des informations sont d'abord enregistrées puis récupérées. Dans le modèle standard d'inférence bayésienne, la mémoire n'intervient pas dans la formation et la mise à jour des croyances. Le décideur reçoit l'information, met à jour ses croyances, et "la croyance postérieure d'hier est égale à la croyance antérieure d'aujourd'hui" (Enke et al., 2019). Pourtant, dès le 18ème siècle, l'économiste et philosophe David Hume (Traité sur Nature Humaine, 1739) avait formulé l'idée selon laquelle la mémoire humaine, de part sa malléabilité, affectait la formation des croyances et donc la capacité de jugement des individus. Selon ses propres mots :

> De même qu'une idée de la mémoire, en perdant sa force et sa vivacité, peut dégénérer d'un tel degré qu'elle est prise pour une idée de l'imagination, de même, de l'autre côté, une idée de l'imagination peut acquérir une telle force et une telle vivacité qu'elle passe pour une idée de la mémoire et en contrefasse les effets sur la croyance et le jugement. (...) Ainsi, il apparaît que la croyance, ou assentiment, qui accompagne toujours la mémoire et les sens, n'est rien que la vivacité des perceptions qu'ils présentent, et que cela seul les distingue de l'imagination. Croire, dans ce cas, c'est sentir une impression immédiate des sens ou une répétition de cette impression dans la mémoire. C'est uniquement la force et la vivacité de la perception qui constituent le premier acte du jugement et qui posent le fondement du raisonnement que nous édifions sur lui quand nous suivons la relation de cause à effet.

Dans cet extrait, Hume décrit les frontières perméables qui existent entre la mémoire et l'imagination. D'après l'auteur, les croyances sont uniquement basées sur la force et la vivacité supérieures des perceptions de la mémoire par rapport à celles de l'imagination. Cependant, comme la vivacité des perceptions peut s'altérer ou

se renforcer au point que la mémoire et l'imagination peuvent se confondre, les croyances sont en fait fondées sur une identification très fragile des perceptions réelles *versus* imaginaires.

Au cours des dernières décennies, l'idée d'une mémoire adaptative, ni parfaite ni stable, mais fondamentalement limitée et dynamique, a acquis une certaine crédibilité en économie. Les économistes ont tenté de modéliser les limitations cognitives de la mémoire et leurs impacts sur la formation des croyances et la prise de décision. En 1991, Dow (1991) est le premier à modéliser un problème de prise de décision d'un agent cherchant à trouver le prix le plus bas d'un objet, en ayant une mémoire limitée des anciens prix. Alors qu'avant le choix du consommateur se limitait aux biens ou aux expériences, Dow (1991) suppose que l'information dont l'agent se souvient est en elle-même un choix. En 1997, Piccione and Rubinstein (1997) modélisent un problème de décision avec mémoire imparfaite. Un décideur fait preuve de mémoire imparfaite si, à un moment donné, il détient de l'information qui est oubliée plus tard. Les auteurs montrent qu'avec une mémoire *parfaite*, le décideur n'a aucune raison de prendre ses décisions à un moment précis (la stratégie optimale définie ex-ante reste optimale pendant son exécution). En revanche, avec une mémoire *imparfaite*, la temporalité des décisions joue un rôle important et le décideur doit définir quand prendre sa décision et dans quelle mesure il peut s'y tenir. Au début des années 2000, Mullainathan (2002) et Bénabou and Tirole (2002) développent simultanément deux modèles avec des hypothèses opposées. Mullainathan (2002) suppose que le décideur est naïf et ignore les imperfections de sa mémoire lorsqu'il fait des inférences. Il montre que les croyances sont influencées non seulement par les informations tirées d'expériences, mais aussi par les souvenirs qu'elles évoquent. Bénabou and Tirole (2002), quant à eux, supposent que le décideur est sophistiqué et a conscience qu'il peut avoir une mémoire sélective lorsqu'il fait des inférences. Plus récemment,

Wilson (2014) a également analysé des problèmes de prise décision avec mémoire imparfaite. Dans son modèle, le décideur est contraint par une capacité limitée de mémorisation. Il ne peut mémoriser qu'un nombre limité de signaux informatifs. L'une des principales implications du modèle est que le décideur ne réagit de façon optimale qu'aux deux signaux les plus extrêmes (et donc les plus informatifs) afin d'éviter de gaspiller des ressources cognitives à mémoriser d'autres signaux moins informatifs. En 2013, Bordalo et al. (2013) ont développé le modèle de la Théorie de l'Attention Dominante. Ce modèle soutient qu'en évaluant plusieurs choix possibles, l'attention du décideur se concentre sur des informations très différentes de celles de son point de référence. Par conséquent, le décideur a tendance à surpondérer ces "caractéristiques surprenantes" lorsqu'il fait des choix. Bien que ce modèle n'inclut aucun mécanisme mnésique, il est à l'origine d'un autre modèle basé sur la mémoire développé cinq ans plus tard par les mêmes auteurs. En 2017, Bordalo et al. (2017) introduisent les deux mécanismes de base de la mémoire, –la répétition et l'associativité–, dans leur modèle de Théorie de l'Attention de la Saillance (Bordalo et al., 2013). La répétition correspond au fait que plus la fréquence à laquelle on se souvient d'un événement est élevée, plus on se souvient de cet événement facilement. L'associativité correspond au fait que la similitude d'un événement passé avec un événement présent permet de se rappeler plus facilement ce dernier événement (Kahana, 2012). Bordalo et al. (2017) constatent que les individus présentent parfois un comportement instable et incohérent dans de nouveaux contextes parce qu'ils fondent leurs choix sur leur mémoire des normes passées, qui ne sont pas nécessairement adaptées à de nouveaux contextes. Considérons par exemple le cas d'un individu se rendant pour la première fois dans un aéroport. Ce dernier a très soif mais, face au prix (qu'il considère) exorbitant des bouteilles d'eau sur place, il décide de ne pas en acheter. L'individu n'achète pas de bouteille d'eau alors même que sa disposition à payer est supérieure au prix demandé car il base son choix sur la mémoire qu'il a du prix "normal" des bouteilles d'eau qu'il

a l'habitude d'acheter en ville. Dans cet exemple, comme dans leur modèle, la mémoire des choix passés détermine (pas toujours de façon efficiente) l'évaluation des choix présents. Dans un cadre similaire de mémoire associative, Bodoh-Creed (2017) modélise la mémoire comme un processus associatif dans lequel la valence de l'humeur déclenche des souvenirs de valence similaire. Dans son modèle, l'humeur est un élément déterministe de la sélectivité des souvenirs. Ainsi, être de bonne humeur (ou d'humeur positive) augmente la probabilité de se rappeler d'un événement positif tandis qu'être de mauvaise humeur (ou d'humeur négative) déclenche le souvenir d'événements négatifs. En appliquant ce modèle aux comportements financiers, il explique les phénomènes de sur-réaction face à l'information et de volatilité du prix des actifs.

Comme nous venons de le voir, la littérature économique modélisant les limitations cognitives de la mémoire et leurs impacts sur la formation des croyances et la prise de décision est riche et variée. Cependant, considérer la mémoire imparfaite n'implique pas nécessairement que les individus manipulent leur mémoire pour répondre à leur besoin de croyances motivées. Un individu peut oublier des informations passées avec une probabilité positive sans pour autant faire preuve de mémoire sélective en fonction de la désirabilité de l'information dont il doit se souvenir. Au contraire, il peut tenter de réprimer, d'oublier et/ou de réinterpréter des informations qui menaçeraient ses croyances ou l'obligeraient à les modifier.

Deux principales raisons sous-tendent le besoin inhérent des humains d'avoir des croyances motivées. Premièrement, les croyances motivées ont une valeur hédonique. Concernant le fonctionnement du monde, le maintien de croyances positives permet de se rassurer sur l'avenir et de maximiser son utilité anticipée (Akerlof and Dickens, 1982; Brunnermeier and Parker, 2005). À cette fin, les individus peuvent préférer maintenir des croyances optimistes favorisant la poursuite

d'une vision de "vie en rose". Concernant l'individu, le maintien de croyances positives peut nourrir l'ego et améliorer la confiance en soi et l'estime de soi (Bénabou and Tirole, 2002; Kőszegi and Rabin, 2006). Cette valeur hédonique des croyances motivées a été étudiée très tôt par Akerlof and Dickens (1982) dans un modèle de réduction de la dissonance cognitive qui repose sur deux propositions. Premièrement, les individus ont des préférences sur leurs croyances. Ces dernières entrent donc directement comme argument dans la fonction d'utilité. Deuxièmement, les individus sont capables de manipuler leurs croyances en choisissant l'information qui est la plus à même de confirmer ce qu'ils ont envie de croire.

La deuxième raison sous-tendant le besoin de croyances motivées est instrumentale (ou fonctionnelle). D'un point de vue motivationnel, le fait d'avoir des croyances positives sur soi ou sur son environnement peut être un puissant moteur pour poursuivre son but, persister face à l'adversité (Bénabou and Tirole, 2006) ou améliorer sa performance[2]. Avoir des croyances optimistes sur soi-même peut aussi aider à convaincre les autres de sa propre valeur et donc être un bon instrument pour atteindre ses objectifs quand ces derniers dépendent des autres (Bénabou and Tirole, 2002). Enfin, les croyances motivées peuvent avoir une valeur instrumentale lorsqu'elles sont utilisées comme une excuse pour justifier un comportement douteux ou irresponsable (Bénabou et al., 2018). Ce dernier point est particulièrement important pour les économistes car un comportement répréhensible peut être coûteux non seulement pour l'individu mais aussi pour l'ensemble de la société. Par exemple, nier ou sous-estimer le changement climatique peut fournir une bonne excuse pour ne pas avoir à s'engager dans un comportement éco-responsable ou

[2]Compte and Postlewaite (2004) ont montré que la valence émotionnelle et la performance étaient positivement corrélées. Ainsi, toute croyance motivée générant ou nourrissant des émotions positives pourrait indirectement agir comme un instrument permettant d'améliorer la performance (Compte and Postlewaite, 2004).

pour continuer à agir de manière préjudiciable.

Très peu de modèles théoriques ont examiné l'utilisation de la mémoire comme stratégie d'auto-tromperie pour répondre à ces besoins, hédoniques et instrumentaux, de croyances motivées. Bénabou and Tirole (2002) ont modélisé la manipulation de la mémoire comme l'équilibre d'un jeu entre les différents "soi" d'un individu, dans lequel ce dernier peut oublier des informations qui peuvent nuire ou menacer sa confiance en lui. Les individus peuvent, avec une certaine probabilité et potentiellement de façon coûteuse, varier la probabilité de se rappeler d'une information (Bénabou and Tirole, 2002). Basé sur ce modèle d'auto-tromperie ("self-deception model") de Bénabou and Tirole (2002), Gottlieb (2014) montre qu'après avoir observé un signal négatif, le décideur fait face à un conflit entre oublier le signal et avoir une meilleure image de soi, ou se le rappeler et prendre une meilleure décision. Lorsqu'il n'y a pas de décision ex post à prendre, le facteur "image de soi" prend le dessus et le décideur se souvient du signal négatif avec une probabilité inférieure au pourcentage réel. Très récemment, Gödker et al. (2019) ont développé un modèle dans lequel les préoccupations liées à l'image de soi déterminent la façon dont l'information est mémorisée par le décideur. Après avoir observé les variations du prix d'un actif dans lequel il a choisi d'investir, l'individu se souvient davantage des variations à la hausse que celles à la baisse. Par conséquent, il pondère de façon asymétrique chaque variation et devient trop optimiste quant à la qualité réelle de l'actif.

Les études expérimentales qui confirment ou réfutent l'utilisation de la mémoire motivée comme mécanisme d'auto-tromperie sont également très limitées. En 1992, Thompson and Loewenstein (1992) ont été les premiers en économie à étudier expérimentalement l'existence d'une mémoire sélective. Ils ont mis en place un jeu de négociation dans lequel deux sujets représentant des partis aux intérêts

divergents devaient négocier avec le parti adverse afin de trouver un accord. En cas d'échec, les deux partis devaient engager des frais. Thompson and Loewenstein (1992) ont constaté que les négociateurs faisaient preuve de mémoire sélective concernant i) les frais encourus en cas d'absence d'accord et ii) les informations échangées lors de la négociation. Les seules autres études expérimentales sur la mémoire motivée en économie, bien plus récentes, sont Li (2013, 2017), Chew et al. (2018), Zimmermann (2019), Carlson et al. (2018) et Gödker et al. (2019).

Chew et al. (2018) montrent qu'après un délai de plusieurs mois, les individus se souviennent de façon asymétrique de leur performance passée dans un test de Quotient Intellectuel (QI). Les individus oublient davantage leurs réponses incorrectes que leurs réponses correctes (mémoire sélective), se souviennent avoir donné des réponses correctes à des questions auxquelles ils n'ont pas été confrontés (illusion), et transforment des réponses incorrectes en réponses correctes (confabulation). En utilisant un test de QI similaire, Zimmermann (2019) trouve également que les individus ont une mémoire biaisée des feedbacks reçus à l'issue d'un test de QI. Précisément, il constate que i) les personnes se rappellent des feedbacks négatifs avec moins d'exactitude que des feedbacks positifs et ii) que les personnes qui ont reçu des feedbacks négatifs déclarent "ne pas se souvenir" plus fréquemment que celles ayant reçu des feedbacks positifs. Dans l'ensemble, Zimmermann (2019) montre que les individus parviennent à supprimer les feedbacks qui menacent leur désir de se considérer comme des personnes intelligentes. Li (2017) teste également si les individus font preuve de mémoire motivée lorsqu'ils se rappellent de leur performance, mais utilise une tâche de saisie de mots au lieu d'un test de QI. Quarante jours après l'exécution de la tâche, les participants doivent se rappeler du nombre d'erreurs qu'ils ont commises et de leur classement. Le design expérimental manipule si les individus doivent prédire leur performance relative ou absolue et s'ils reçoivent ou non un feedback. Li (2017) trouve que le fait d'avoir

à prédire sa performance et de recevoir un feedback élimine les biais de mémoire. Gödker et al. (2019) trouvent également des preuves expérimentales de l'existence d'une mémoire motivée, mais dans des décisions financières et non concernant une performance individuelle. Dans leur expérience, les sujets choisissent d'investir soit dans un actif risqué, soit dans un actif non-risqué, puis observent l'évolution du prix de l'actif. Par la suite (soit immédiatement après, soit une semaine après selon le traitement), les sujets doivent se rappeler des différents prix observés de l'actif. Gödker et al. (2019) constatent que les sujets qui ont investi dans l'actif risqué se souviennent mieux des gains réalisés que des pertes encourues, forment des croyances optimistes sur l'évolution future du prix de l'actif, et sont davantage susceptibles de réinvestir en bourse que les individus ayant initiallement choisi un actif non-risqué.

L'image de soi ne dépend pas que de son intelligence et/ou de ses performances individuelles. En tant qu'animaux sociaux, la demande d'une image positive de soi est aussi fortement liée au désir de paraître pro-social, non seulement aux yeux des autres (Bénabou and Tirole, 2002; Battigalli and Dufwenberg, 2007), mais aussi à ses propres yeux (Ariely et al., 2009; Grossman and Van Der Weele, 2017). Tandis que certains individus sont profondément animés par des motivations altruistes, d'autres investissent dans de bonnes actions pour maintenir une bonne image d'eux-mêmes, apportant ainsi la preuve qu'être pro-social joue un rôle important dans la construction de l'image de soi (Bénabou and Tirole, 2006). Pourtant, certaines situations présentent parfois un arbitrage entre deux choix possibles : favoriser les autres à ses propres dépens, ou se faire passer en premier au détriment des autres. Lorsque la balance penche en faveur du second choix, les individus peuvent s'engager dans des actions qui nuisent aux autres, contrariant ainsi leur demande d'image pro-sociale et pouvant potentiellement créer des incohérences avec leurs propres préférences (Banaji and Bhaskar, 2000; Banaji

et al., 2004; Chugh et al., 2005; Tenbrunsel et al., 2010). Une façon de rétablir ces contradictions ou incohérences est d'avoir une mémoire motivée. Li (2013) est le premier à avoir étudié la manipulation de la mémoire dans les interactions sociales. Après avoir joué un jeu de confiance ("trust game"), les deux types de joueurs (les "trustees" et les "trustors") doivent se souvenir des décisions prises au cours du jeu. Li (2013) ne trouve aucune preuve de mémoire sélective de la part des trustees selon qu'ils ont réciproqué ou trahi le trustor. Seuls les trustors font preuve de mémoire sélective en se souvenant moins bien de leurs décisions lorsqu'ils ont été trahis que lorsqu'ils ont été récompensés pour leur confiance. En revanche, Carlson et al. (2018) trouvent, dans un jeu du dictateur, qu'il est plus probable que les dictateurs se souviennent incorrectement du montant alloué au joueur passif lorsque la décision qu'ils ont prise contredit leur vision personnelle de la justice et de l'équité.

Cette thèse étudie l'existence et la force de mémoire motivée dans trois contextes économiquement pertinents : les préférences sociales, la performance individuelle et les décisions malhonnêtes. Dans chaque contexte, elle fournit un nouvel ensemble de preuves montrant que les individus manipulent leur mémoire pour soutenir leur besoin d'avoir des croyances qui leur sont confortables ou favorables. Elle explore deux potentiels déterminants de la mémoire motivée : un déterminant hédonique lorsque la manipulation de la mémoire sert à améliorer l'image de soi, et un déterminant instrumental lorsqu'elle sert d'excuse pour justifier ses décisions futures. Chaque chapitre présenté ici se concentre sur la compréhension d'un déterminant de la mémoire motivée dans un contexte spécifique. Plus précisément:

- Le chapitre 1 examine si les gens ont une mémoire biaisée de leurs interactions avec autrui. Les individus oublient-ils les conséquences de leurs actes sur les autres ? Si oui, cela dépend-il de la nature (par exemple, égoïste ou altruiste) de leurs actions ? Nos résultats confirment la sélectivité des souvenirs dans les interactions sociales. Les individus se souviennent davantage de leurs

décisions altruistes que de leurs décisions égoïstes. En revanche, nous ne trouvons pas preuve évidente de biais d'erreurs de mémoire.

- Le chapitre 2 distingue deux forces qui ont été proposées pour expliquer l'existence de biais de mémoire concernant la performance individuelle. Pourquoi les individus se souviennent mieux des feedbacks positifs que des feedbacks négatifs ? Alors que l'hypothèse d'auto-amélioration affirme que les individus se souviennent mieux des feedbacks positifs pour améliorer leur image de soi, la mémoire associative assure que les individus se souviennent mieux de l'information qui est en accord avec leur humeur. Nous proposons un environnement contrôlé dans lequel les deux théories prédisent des résultats différents. Nos résultats supportent l'existence et la dominance relative d'une mémoire d'auto-amélioration de l'image de soi par rapport à celle d'une mémoire congruente à l'humeur.

Les chapitres 1 et 2 se concentrent sur le cas où les souvenirs ont une valeur purement hédonique. Le décideur ne prend pas de décision ex-post, et la seule raison motivant la manipulation de la mémoire est l'amélioration de son image. Le chapitre 3 examine le cas où les souvenirs ont également une valeur instrumentale.

- Le chapitre 3 examine le rôle respectif des motifs hédoniques et instrumentaux dans la mémoire motivée, avec une application au domaine de l'éthique. Les individus ont-ils une mémoire motivée de leurs comportements malhonnêtes ? Si oui, est-ce dû à des motifs purement hédoniques et/ou à des raisons stratégiques ? Nous trouvons que les considérations hédoniques seules ne sont pas suffisantes pour déclencher une manipulation de la mémoire. En revanche, quand l'oubli sert d'excuse pour justifier leurs décisions futures, les individus manipulent leur mémoire.

Les sections suivantes présentent chaque chapitre de cette thèse et leurs contributions originales à la littérature existante.

Chapitre 1: Mémoire Motivée dans des Jeux du Dictateur

Le Chapitre 1 vise à comprendre si et dans quelle mesure les individus manipulent leur mémoire pour soutenir leur demande d'image pro-sociale. Comme Li (2013), nous visons à identifier l'existence d'une mémoire motivée et à étudier la sélectivité des souvenirs dans les interactions sociales. Contrairement à Li (2013), nous sommes en mesure i) d'étudier à la fois la mémoire sélective (lorsque la probabilité de se souvenir d'un acte désirable est supérieure à celle de se souvenir d'un acte indésirable) *et* le biais des erreurs de mémoire (qui se définit par la direction et la magnitude des erreurs), et ii) de déterminer un effet causal de responsabilité de la décision sur la mémoire sélective. Nous avons conçu une expérience en laboratoire où les participants doivent jouer à une série de jeux de dictateurs binaires puis, après avoir réalisé une tâche visant à distraire leur attention, se rappeler des montants attribués au joueur passif. Nous avons introduit quatre traitements dans lesquels nous manipulons la responsabilité des dictateurs pour le montant alloué au joueur passif (soit le dictateur choisit le montant, soit le montant est choisi au hasard par un ordinateur) et la présence d'incitations monétaires pour un souvenir correct.

Nos résultats montrent que les individus ont une mémoire sélective. Premièrement, lorsque les dictateurs sont responsables du montant alloué au joueur passif, le pourcentage de souvenirs corrects est plus élevé après avoir choisi l'option altruiste qu'après avoir choisi l'option égoïste. Cela n'est pas le cas lorsque le montant du joueur passif est sélectionné au hasard par l'ordinateur. Deuxièmement, la présence d'incitations monétaires pour des souvenirs corrects augmente le pour-

centage de souvenirs corrects chez les dictateurs, mais seulement lorsqu'ils ont choisi l'option altruiste, pas lorsqu'ils ont choisi l'option égoïste. Cela suggère que les individus n'oublient pas complètement leurs décisions passées, mais lorsqu'ils sont incités financièrement à fournir un effort de mémoire, ils consacrent cet effort à recouvrer la mémoire d'informations valorisantes plutôt que dévalorisantes. Enfin, lorsqu'on demande aux dictateurs de se souvenir non pas du montant alloué au joueur passif mais de leur propre montant, ils sont également moins susceptibles de s'en souvenir après avoir choisi l'option égoïste que l'option altruiste. Ce résultat montre que les souvenirs sélectifs ne sont pas motivés par une attention accrue accordée au montant du joueur passif par des dictateurs pro-sociaux.

En revanche, nous ne trouvons pas de preuves claires de biais d'erreurs de mémoire. Les dictateurs sont plus susceptibles de sur-estimer que de sous-estimer le montant alloué au joueur passif après avoir choisi l'option égoïste plutôt que l'option altruiste, mais la même asymétrie se retrouve lorsque le montant alloué au joueur passif est choisi de façon aléatoire par le programme informatique. De plus, la magnitude des erreurs de mémoire des dictateurs est semblable quelque soit le degré de pro-socialité de la décision, et indépendamment du fait que le dictateur en soit responsable ou pas. Même si la majorité des individus préfèrent probablement se considérer comme généreux plutôt qu'égoïstes, une explication possible de l'absence de biais d'erreurs de mémoire est que la dissonance entre la prise de décisions égoïstes quand une alternative pro-sociale est disponible et le maintien d'une image positive de soi peut ne pas être assez forte pour générer un conflit interne. L'étude de la mémoire motivée dans le domaine de la moralité et de l'éthique, où les impératifs catégoriques (Kant, 1785) et les normes injonctives sont plus importants, pourrait générer un besoin accru de mémoire biaisée. De plus, dans les études qui ont été présentées jusqu'à présent, les individus ne pouvaient manipuler leur mémoire que pour des raisons hédoniques, c'est-à-dire pour

améliorer leur image d'eux-mêmes et se percevoir comme plus intelligents et/ou généreux. De telles considérations hédoniques, à elles seules, pourraient ne pas être suffisantes pour déclencher des erreurs de mémoire biaisées. L'introduction de raisons instrumentales, c'est-à-dire de situations dans lesquelles la manipulation de la mémoire peut servir d'excuses ou de justifications pour des décisions *futures*, pourrait permettre de mieux identifier et comprendre l'émergence des erreurs de mémoire biaisées et leur rôle dans le processus de prise de décision. L'exploration de ces deux possibilités est l'objet du troisième chapitre de cette thèse. Le chapitre 2 continue d'explorer les distorsions de la mémoire pour des raisons hédoniques, mais dans le domaine de la performance individuelle plutôt que dans celui des interactions sociales. Il étudie les mécanismes sous-jacents de la mémoire motivée.

Chapitre 2: Les Biais de Mémoire sont-ils dépendants de l'Humeur ou de l'Image de Soi?

Les feedbacks que nous recevons sont l'une des principales sources d'information sur nous-même : ils peuvent aider les personnes à combler un manque d'information, à actualiser leurs croyances, à s'améliorer, à faire de meilleurs choix et ainsi obtenir de meilleurs résultats. Bien que plusieurs études aient mis en évidence l'existence d'une asymétrie dans le rappel de feedbacks concernant une performance individuelle passée, les raisons qui sous-tendent cette asymétrie ne sont pas claires. Dans la littérature, deux hypothèses ont été adoptées pour expliquer cette asymétrie des souvenirs. D'une part, la mémoire auto-améliorante soutient que les individus se souviennent mieux des feedbacks positifs que des feedbacks négatifs afin de maintenir une bonne image d'eux-mêmes (effet d'auto-amélioration). D'autre part, la mémoire associative assure que les biais de mémoire sont dûs une à accessibilité accrue de l'information positive et à une accessibilité atténuée de l'information négative lorsque les gens sont d'humeur non-négative (effet de congruence de l'humeur).

Bien que les deux principes ne soient pas mutuellement exclusifs et que la plupart des preuves existantes soient compatibles avec les deux théories, il est essentiel de comprendre la nature des mécanismes en actions. Cela permettrait de prévenir certaines décisions économiques sous-optimales et d'élaborer des politiques visant à atténuer ou à éliminer les comportements de sur-confiance, si nécessaire. En effet, alors que les modèles d'auto-amélioration supposent un contrôle méta-cognitif de l'individu et suggèrent ainsi l'importance de corriger les croyances ex ante sur ce qui est utile ou non pour l'individu, les modèles de congruence de l'humeur supposent un processus heuristique sans intentionnalité et considèrent les biais de mémoire comme étant un collatéral d'état affectifs positifs.

Le Chapitre 2 a pour but de démêler ces deux forces qui ont été proposées comme explication possible des biais de mémoire concernant la performance individuelle. Pour identifier et démêler ces deux mécanismes, nous avons mis en place une expérience en laboratoire où les deux théories, la mémoire auto-améliorante et la mémoire congruente à l'humeur, offrent des prédictions divergentes. Dans notre expérience, basée sur le design de Zimmermann (2019), les sujets effectuent un test de QI puis reçoivent un feedback bruité sur leur performance relative par rapport à leurs pairs. Un mois plus tard, ils reviennent au laboratoire et doivent se rappeler des feedbacks reçus un mois auparavant. Avant de se souvenir, nous intervenons ou non sur leur humeur, en utilisant la procédure d'Andrade et al. (2015). Le laboratoire offre un environnement contrôlé où la précision du souvenir peut être soigneusement évaluée et l'humeur manipulée de façon exogène.

Nos résultats supportent l'existence d'une mémoire d'auto-amélioration. Premièrement, les personnes se souviennent mieux des feedbacks positifs que des feedbacks négatifs. Deuxièmement, lorsqu'ils ne se souviennent pas correctement du feedback reçu un mois auparavant, les individus présentent des erreurs de mé-

moire optimistes. Cela signifie qu'ils surestiment le nombre de feedbacks positifs qu'ils ont reçu. Ces résultats répliquent les résultats de Zimmermann (2019). En revanche, bien que notre manipulation de l'humeur s'avère efficace pour induire l'état affectif souhaité, nous ne trouvons pas de preuve claire de mémoire congruente à l'humeur. Les personnes ne se souviennent pas mieux des feedbacks lorsque ces derniers ont une valence congruente à celle de l'humeur induite en laboratoire. Ces résultats confirment l'effet de la mémoire d'auto-amélioration comme moteur des biais de souvenirs asymétriques sur la performance individuelle, mais ne montrent aucun rôle de la congruence de l'humeur.

Ces résultats démontrent l'importance des facteurs motivationnels plutôt qu'affectifs dans la formation de croyances optimistes sur soi. La prédominance relative d'une mémoire motivée par l'auto-amélioration a des répercussions directes sur les politiques visant à atténuer ou à éliminer les biais de jugement. Dans la mesure où les individus déforment surtout leur mémoire parce qu'ils considèrent les feedbacks négatifs comme potentiellement nuisibles, l'élimination de l'aversion ex ante pour ces feedbacks négatifs pourrait permettre d'atténuer ces biais de mémoire pouvant potentiellement conduire à une sur-confiance en soi pas toujours souhaitable (Bénabou and Tirole, 2002).

Chapitre 3: Mémoire Motivée des Comportements Malhonnêtes

Plusieurs études en économie ont identifié différentes stratégies utilisées par les individus pour maintenir leurs valeurs morales tout en agissant de façon contraire à l'éthique. Les individus évitent de connaître les conséquences de leurs actes (Feiler, 2014; Grossman and Van Der Weele, 2017), exploitent l'incertitude sur la norme sociale en vigueur pour justifier leur propre décision de mentir (Bicchieri et al., 2019), prétendent avoir changé (Stanley et al., 2017), rejettent la faute sur autrui

(Bartling and Fischbacher, 2011; Oexl and Grossman, 2013), contre-balancent leur comportements (im)moraux dans le temps (Ploner and Regner, 2013; Gneezy et al., 2014; Cojoc and Stoian, 2014) ou trouvent des excuses consistant, par exemple, à minimiser les externalités négatives et/ou réinterpréter les circonstances de leurs actions (Bénabou et al., 2018). D'autant que nous le sachions, aucune étude n'a à ce jour exploré la manipulation de la mémoire comme un autre moyen de maintenir ses valeurs morales tout en agissant de façon malhonnête.

Le Chapitre 3 étudie la présence de mémoire motivée pour des raisons hédoniques mais aussi instrumentales, dans un contexte de prise de décision (non) éthique. Notre étude contribue à deux littératures. Premièrement, elle contribue à mieux comprendre le raisonnement moral lorsqu'on est confronté à une occasion de mal se comporter. Malgré la littérature abondante sur la malhonnêteté qui s'est développée au cours de la dernière décennie (pour les surveys, voir notamment Rosenbaum et al., 2014; Irlenbusch and Villeval, 2015; Jacobsen et al., 2018; Abeler et al., 2011), il n'existe aucune étude économique sur la manipulation de la mémoire comme mécanisme permettant de soutenir une image morale de soi tout en agissant de façon malhonnête. Deuxièmement, elle contribue à la littérature économique récente mais croissante sur la mémoire motivée. Alors que les études existantes (Li, 2013, 2017; Chew et al., 2018; Zimmermann, 2019; Carlson et al., 2018) étudient la mémoire motivée dans le domaine de la performance individuelle ou des préférences sociales, nous nous concentrons sur la mémoire des décisions malhonnêtes ou contraires à l'éthique. De plus, nous explorons non seulement si les individus oublient leurs comportements malhonnêtes passés pour soutenir leur désir d'image morale de soi, mais aussi s'ils manipulent leur mémoire comme une excuse pour ne pas avoir à s'engager dans un comportement moralement responsable mais coûteux. Ainsi, nous proposons le premier test expérimental de l'impact des décisions *anticipées* sur la manipulation de la mémoire. Nous étudions les

biais de la mémoire non seulement comme une *conséquence* d'un comportement contraire à l'éthique, mais aussi comme un *instrument* pour justifier un futur comportement potentiellement non éthique.

Pour étudier i) si les individus oublient leurs comportements malhonnêtes passés et ii) s'ils utilisent leur mémoire comme instrument pour justifier leurs décisions futures, nous avons mené une expérience en ligne dans laquelle les participants jouent à un jeu de hasard répété. Dans ce jeu, les individus observent le résultat d'un tirage au sort et peuvent décider de mentir sur ce résultat afin de maximiser leurs gains. Trois semaines plus tard, les individus doivent se souvenir des montants qu'ils ont reporté lors de la première session. Selon les traitements, nous varions i) si les individus peuvent ou non tricher en session 1 et ii) si seules des raisons hédoniques ou à la fois hédoniques et stratégiques peuvent pousser les individus à oublier leurs comportements passés.

Nos résultats montrent que lorsque les biais de mémoire n'ont qu'une valeur hédonique –c'est-à-dire améliorer ou conserver une bonne image de soi–, les personnes malhonnêtes ne se souviennent pas moins bien de leurs décisions passées que les personnes qui ne pouvaient pas tricher en session 1. Ainsi, dans notre contexte, les considérations hédoniques ne sont pas suffisantes pour déclencher une manipulation significative de la mémoire. En revanche, lorsque les biais de mémoire ont une valeur instrumentale –c'est-à-dire lorsque les individus sont informés, *avant* de se souvenir des montants reportés, qu'ils auront une décision future à prendre–, les individus malhonnêtes se souviennent de leurs comportements passés avec moins de précision que lorsqu'ils savent qu'ils n'auront aucune décision à prendre. Ce résultat suggère que les personnes utilisent leur mémoire comme une excuse pour justifier leurs décisions futures. Il confirme que la mémoire est impliquée dans les différentes stratégies que les individus utilisent pour motiver leurs croyances sur

eux-mêmes et justifie qu'ils peuvent se comporter immoralement tout en gardant une opinion positive d'eux-mêmes.

***

L'utilisation d'expériences en laboratoire pour étudier les déterminants comportementaux des biais de mémoire présente de nombreux avantages. Tout d'abord, les expériences en laboratoire permettent d'observer la mémoire d'informations (décisions et/ou feedback) directement induites au sein laboratoire et ainsi de mesurer précisément les erreurs de mémoire. Cela n'est pas nécessairement le cas dans des études portant sur la mémoire auto-déclarée ou autobiographique. Deuxièmement, les expériences en laboratoire permettent de manipuler de manière exogène la nature de l'information dont il faudra par la suite se souvenir et donc de tester la récupération de la mémoire en fonction du degré de désirabilité de l'information. Ce point est particulièrement important lorsqu'il s'agit d'étudier la manipulation de la mémoire comme stratégie utilisée par les individus pour soutenir leur désir de croyances motivées. Enfin, les expériences de laboratoire permettent à l'expérimentaliste de contrôler la durée de l'expérience entre les différentes phases d'encodage et de remémoration, et ainsi de distinguer l'effet du temps de l'effet de la motivation sur les mécanismes mnésiques.

L'étude de la mémoire humaine via des expériences de laboratoire présente cependant certaines limites spécifiques. Premièrement, dans les expériences en laboratoire, il est de connaissance commune que l'expérimentaliste détient l'information dont les participants doivent se rappeler. Une conséquence directe est que les participants peuvent être confrontés à un arbitrage entre deux stratégies de renforcement de l'image de soi lorsqu'ils doivent se souvenir de leurs comportements passés. D'un côté, oublier une action ou un comportement dont on est peu fier permet

d'améliorer son image de soi. De l'autre, cela peut également l'endommager car oublier envoie un signal (négatif) indiquant que sa mémoire ne fonctionne pas correctement. Le résultat d'un tel arbitrage dépend de la nature même de l'individu. Deuxièmement, dans la vie réelle, les événements vécus sont rarement enregistrés et donc non vérifiables. Dans de tels environnements incontrôlés, la manipulation de la mémoire peut survenir plus librement que dans un laboratoire où l'individu peut être confronté à des faits tangibles et ainsi connaître des coûts intrinsèques plus élevés de manipulation de la mémoire. De plus, toutes les études économiques qui ont exploré la mémoire en laboratoire rémunèrent les souvenirs corrects. Bien que les incitations monétaires soient cruciales pour tester l'existence et la force des biais de mémoire, leur présence peut aboutir à une mesure conservatrice de ces derniers. Dans la vie réelle, les incitations monétaires peuvent parfois être alignées avec l'oubli ou la distorsion des souvenirs. Pour donner un example provocateur, nous pouvons par exemple penser à un individu qui oublie (probablement de façon volontaire mais éventuellement sans le faire exprès) son portefeuille en allant au restaurant. De façon moins anecdotique, le fait de présenter des troubles de la mémoire peut conduire à des réductions de peine de la part des tribunaux de grande instance lors de procès faisant suite à des actes criminels (Cima et al., 2002).

Contents

List of Tables

List of Figures

Foreword

Memory speaks to everyone. We lose it, we recover it, we muscle it. Although the use of a single term might seem to suggest that memory works as a unitary system, the scientific investigations on memory that started about 100 years ago have come to show that memory is not one but many systems (Baddeley, 1997). The systems range in nature (episodic memory, procedural memory, semantic memory, etc.), in encoding (visual memory, olfactory memory, sensitive memory), in storage duration and capacity (from few seconds to long-term memory), and in scale (since memory is not only individual but can be collective).

Memory is dynamic, flexible, multi-dimensional, and thereby complex. This is why most specialists agree to talk about *memories* rather than *memory*. If some of them outside the economic sphere were to read the studies presented here, as probably any other study in economics on memory, they would probably be mocking (if not outraged) by the way we consider memory as a black box. In our framework, individuals receive information (from their own actions or from external feedback) and are later asked to retrieve it. We readily acknowledge that such a framework is restrictive in many regards. Notably, beyond ignoring the brain mechanisms (particularly the neural networks and cellular plasticity allowing the different processes of consolidation and re-consolidation of memory over time, which is of main interest for neuroscientists); it focuses on a very specific memory system −episodic

memory−, among many others that have been identified (see Figure 2). The use of the term *memory* in economics and throughout this thesis always refers to the episodic system of memory. It corresponds to the memory of autobiographical events, such as remembering your first day at school or that day you received a biting rejection from a high-ranked journal. It is different from semantic memory that refers to more general knowledge that we have accumulated throughout our lives, such as knowing the capital of France or the ranking of the top five journals. These two systems belong to the declarative memory, which refers to the conscious, intentional recollection of past events and facts (Ullman, 2004). Other types of memory systems are relying on implicit memory that, by contrast, is acquired and used unconsciously (Schacter, 1987).

This restrictive vision of memory in economics being acknowledged, we are convinced that introducing memory into the economic discipline remains extremely important. The economist does not necessarily need to know what is in the black box of memory and how it is processed. What is important, however, is to know how and to what extent memory intervenes in the input-decision process. Memory of past experiences is one of the main sources of information about ourselves and the world surrounding us. Exploring how individuals retrieve such information is important to better understand how they form and update their beliefs. It changes the understanding of how people learn from experiences to make decisions. Studying (asymmetric) recall of information also permits to better apprehend the emergence of inaccurate statements about oneself, such as overconfidence, with major implications on the quality of choices. Also, because "anything that distort individuals' ability to recall an event will distort their probability assessments" (Hammond et al., 2006), exploring the behavioral determinants of memory biases may grant a better understanding of how individuals form anticipations and ex-

pectations.

Introducing memory into the process of individuals' decision-making in economics highlights a new and serious candidate to explore the dynamics of motivated beliefs. It offers an alternative channel to explain suboptimal decisions when individuals do not adjust for the fallibility of their memory when making choices. As philosophers, psychologists, sociologists or neuroscientists may explore the concepts of identity, ethics or communication in very different ways, economists should not be afraid of investigating human memory in a restrictive or simplified way, as long as it allows them to better explain how people make decisions in a world with scarce resources, both material and cognitive.

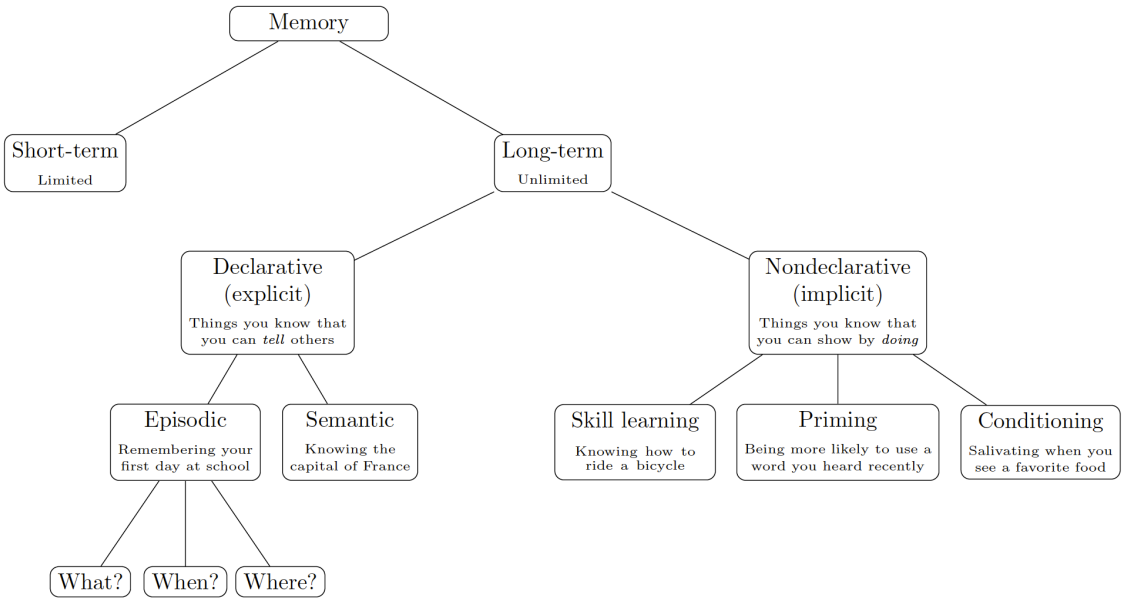


Figure 2: Memory Systems (Adapted from Raslau et al. 2014)

General Introduction

The desire to hold positive beliefs about oneself and the world surrounding us is a fundamental feature of humans. Individuals cherish subjective beliefs and invest huge amounts of effort to protect them. For instance, parents like to think that their children are particularly advanced in their development or abilities, job candidates like to think that they are more able and competent than other candidates, people on diet easily believe that this (not so) occasional ice cream will not really make a difference, some travellers minimize the environmental footprint of their behavior, etc. As in these everyday life examples, often beliefs are not formed and updated in a neutral manner but self-servingly in response to psychological or emotional needs. Individuals collect and interpret evidence in a fashion that supports what they are motivated to believe. Beliefs are not just held, but *motivated* (Bénabou, 2015).

This conception of beliefs as an "asset that people invest in" (Bénabou and Tirole, 2011) has profoundly challenged the vision of the *homo oeconomicus* (Mill, 1874) that is supposed to form and adjust his beliefs in a rational response to new information. In fact, as Schelling (1987) described "the mind as a consuming organ", individuals may obtain satisfaction not only from the consumption of goods or experiences but also from the consumption of beliefs. As a response, behavioral economists have introduced the concept of "belief-based utility" in which individu-

als can derive direct utility from motivated beliefs, even if the latter are inaccurate and can conflict with optimal decision-making.

Individuals, however, do not simply believe what they want to believe. As it would be hard for a starting PhD student to convince himself that he is currently the best economist in the world, a shareholder will never be certain that the value of his portfolio will never decrease, even temporarily. Beliefs have to be plausibly motivated to be held. To satisfy this demand for motivated beliefs, individuals thus need to develop strategies through which they can defend themselves and reason against potentially threatening evidence.[3] The puzzling question is: how? Economists have been studying this issue for years, and several types of strategies have been explored.

Ex ante, evidence shows that individuals sometimes prefer the "bliss of ignorance" rather than the "power of knowledge". Individuals avoid or selectively acquire information to sustain desirable beliefs, even when it is free and it could improve decision making (Golman et al., 2017). Examples go from investors avoiding looking at their portfolios when the market is down (Karlsson et al., 2009) to individuals eschewing free medical tests (Oster et al., 2013). Information avoidance produces uncertainty which impacts individuals' decisions. Numerous studies have shown that individuals engage in willful or strategic ignorance of inconvenient information as an excuse to act self-interestedly. For instance, Dana et al. (2007) have shown that dictators use uncertainty on the consequences of their decisions as a moral wiggle room to act more selfishly than when they are informed, because ignorance protects their self-image (see also Grossman, 2014; Kajackaite, 2015;

[3]It is assumed here that information available to individuals is not manipulated *per se*. This is a restrictive representation of the reality since it seems plausible to think that institutions such as governments, medias, politician parties and/or enterprises may sometimes have interest to provide *ex-ante* distorted or fake information.

Grossman and Van Der Weele, 2017). Individuals also give less to charities under risk (Exley, 2015) and uncertainty (Garcia et al., 2018) than when the outcome of their decisions is certain. Interestingly, willful ignorance matches a supply of ignorance. In a laboratory experiment, Shalvi et al. (2019) have shown that not only some decision-makers avoid advisers who transmit inconvenient information (demand side), but a majority advisers themselves suppress inconvenient information transmitted to the decision-makers (supply side).

Ex post, accuracy is not always the main goal underlying beliefs formation either. Individuals often weight information asymmetrically and update beliefs in a non-bayesian way (Rabin and Schrag, 1999). For instance, Eil and Rao (2011) and Mobius et al. (2011) have shown that subjects' belief updating of good news about their IQ (and beauty in Eil and Rao, 2011) conformed more to Bayes' Rule that belief updating of bad news. Also using an IQ test, Zimmermann (2019) found that beliefs are adjusted in the appropriate direction when elicited directly after feedback is given, but are asymmetrically updated when elicited one month after receiving feedback. Precisely, beliefs of participants who received positive feedback remained adjusted upwards while beliefs of participants who received negative feedback tended to return to their priors. Beyond distorting information, individuals can defend themselves against threatening evidence by shifting the blame onto someone else for their actions (Bartling and Fischbacher, 2011; Oexl and Grossman, 2013), dilute their responsibility (Foerster and Van der Weele, 2018) or use narratives consisting in, for example, downplaying the externalities, and/or reinterpret the circumstances of their actions (Bénabou et al., 2018; Foerster and van der Weele, 2018).

While *ex ante* willful avoidance and *ex post* biased updating or distortion of information have been widely studied by experimental economists, another strategy available to individuals has long been disregarded. Even when information has been received and updated, individuals can ultimately exhibit *forgetting* or *biased memory* of information. People may fail to retrieve undesirable information and, reversely, strive to remember desirable one and even fake memories. In other words, time could give individuals the wiggle room to forget or distort what they would rather not remember. This disregard for memory by economists seems surprising when considering that most of the beliefs that underlie our decisions are drawn from memory (Tranel et al., 1994; Bechara et al., 1998). In fact, our memory of past events is, if not the first, one of the main sources of information. Introducing memory into the process of individuals' decision-making thus seems of crucial importance. It highlights a new and serious candidate to explore the dynamics of motivated beliefs. This is of importance for economists since memory biases may lead to inaccurate statements about oneself and one's environment, with major implications on the quality of choices.

This thesis investigates whether individuals use their memory as a self-deceptive strategy to sustain their desire for motivated beliefs. It tests experimentally the existence and strength of memory manipulation in economically relevant contexts such as social interactions, individual performance and unethical decisions. It provides a novel body of evidence that even when information has been received and updated, another strategy available to individuals to manipulate their beliefs about themselves is indeed to forget such information.

> I have done this, says my memory. I cannot have done that, says my pride, remaining inexorable. Finally, memory yields.
>
> Nietzsche, 1886, *Beyond Good and Evil*

Memory has long been seen as a storehouse into which discrete items of information are initially deposited and later retrieved. In the standard model of Bayesian updating, memory does not intervenes in the belief formation and updating process. The decision-maker updates his beliefs upon receipt of information and "yesterday's posterior equals today's priors" (Enke et al., 2019). Yet, in the 18th century, the idea that humans' memory malleability affected beliefs had been formulated by the economist and philosopher David Hume in his Treatise of Human Nature (1739). In his own words:

> As an idea of the memory, by losing its force and vivacity, may degenerate to such a degree, as to be taken for an idea of the imagination; so on the other hand an idea of the imagination may acquire such a force and vivacity, as to pass for an idea of the memory, and counterfeit its effects on the belief and judgment. (...) Thus it appears, that the belief or assent, which always attends the memory and senses, is nothing but the vivacity of those perceptions they present; and that this alone distinguishes them from the imagination. To believe is in this case to feel an immediate impression of the senses, or a repetition of that impression in the memory. It is merely the force and liveliness of the perception, which constitutes the first act of the judgment, and lays the foundation of that reasoning, which we build upon it, when we trace the relation of cause and effect.

Hume's essential point is to recognize the permeable boundaries between memory and imagination and its impacts on beliefs. Hume points out that beliefs are solely based on the superior force and vivacity of perceptions from memory

over imagination. However, since the vivacity of perceptions can degenerate or strengthen to a point where memory and imagination may confound, beliefs are in fact based on a very fragile identification of actual *versus* fancy perceptions.

Over the last decades, the idea of an adaptive memory, neither perfect nor stable, but fundamentally limited and dynamic, has gained some credence in economics. Economists have attempted at modelling cognitive limitations in recalls and their impact on belief formation and decision-making. In 1991, Dow (1991) is the first to model the decision problem of an agent searching for a low price, with a limited memory of past prices. While before the consumer's choice was restricted to goods or experiences, he assumes that the information that is remembered is *itself* a choice variable. In 1997, Piccione and Rubinstein (1997) model a decision problem with imperfect recall. A decision-maker exhibits imperfect recall if, at a point of time, he holds information which is forgotten later on. The authors show that while under perfect recall the decision-maker has no reason to make his decision at a particular point of time (the ex-ante defined optimal strategy remains optimal during its execution), under imperfect recall the timing of decision plays an important role and the decision-maker needs to define when to make his decision and to what extent he can commit to it. In the early 2000s, Mullainathan (2002) and Bénabou and Tirole (2002) developed simultaneously two theoretical models with contrasting hypotheses. Mullainathan (2002) assumes that the decision-maker is naive and ignores memory imperfections when making inferences. He shows that beliefs are affected not only by information drawn from experience but also by the memories they evoke. On the contrary, Bénabou and Tirole (2002) assume that the decision-maker is sophisticated and realizes that he may select memories when making inferences. More recently, Wilson (2014) has also analyzed decision problems with imperfect recall. In his model, the decision-

maker is constrained by a finite memory capacity. He cannot memorize all but only a limited number of informative signals. Under such limited memory, the decision-maker optimally reacts only to the two most extreme signal realizations, which prevents him from wasting limited memory resources on less informative signals. In 2013, Bordalo et al. (2013) have developed the Salience Theory of Attention model. This model holds that, when evaluating a choice option, the decision-maker's attention focuses on the information that differs the most from his reference point. As a consequence, the decision-maker tends to overweight the "surprising features" when making choices. While this model does not consider memory mechanisms, it is a the origin of another memory-based model developed five years later by the same authors. In 2017, Bordalo et al. (2017) introduce the two baseline mechanisms of memory, –rehearsal and associativeness–, into their Salience Theory of Attention model (Bordalo et al., 2013). Rehearsal corresponds to the fact that the higher frequency to which an event is remembered makes it easier to remember again. Associativeness corresponds to the fact that the similarity of a past event to a current event makes this latter event easier to recall (Kahana, 2012). Bordalo et al. (2017) find that individuals sometimes exhibit unstable and inconsistent behavior in new contexts because they base their choices on their memory of past norms, which are not necessarily adapted for new contexts. For instance, a first-time traveler at the airport may not buy an expensive bottle of water even if he is extremely thirsty, just because he has in memory a norm of "low price" for bottles of water he brought downtown. In that example, as in their model, memory of previous choices shapes (not always consistently) the evaluation of current choices. In a similar associative recall framework, Bodoh-Creed (2017) models memory as an associative process where the current mood (or affective state) is a relevant cue for retrieval. In a dynamic setting, he incorporates mood as a deterministic element of selective recall under the assumption that positive

mood increases the probability to recall a positive item while negative mood triggers recollection of negative items. By applying the model to financial behavior, he draws predictions on information overreaction and asset price volatility.

As just reviewed, the theoretical economic literature modeling cognitive limitations in recalls and their impact on belief formation and decision-making is now substantial. However, considering imperfect memory does not necessarily imply that individuals use memory manipulation as a self-deceptive strategy. For instance, individuals may forget past information with some positive probability without exhibiting asymmetries in recalls depending on the desirability of the information to be retrieved. By contrast, individuals may try to repress, forget and/or reinterpret information to sustain their desire for motivated beliefs.

Why do individuals need motivated beliefs? Two main reasons underlying the demand for motivated beliefs have been identified in the literature (mainly theoretical). First, motivated beliefs have an *hedonic* value. Holding positive beliefs about oneself and the world surrounding us is just pleasant *per se*. Maintaining positive beliefs about how the world works may reassure oneself about the future, maximize anticipatory utility and protect one's "rosy view" of the world (Akerlof and Dickens, 1982; Brunnermeier and Parker, 2005). Maintaining positive beliefs about the self may feed the ego, enhance self-confidence and self-esteem, and make oneself feel better (Bénabou and Tirole, 2002; Kőszegi and Rabin, 2006). Maintaining positive beliefs may also prevent from bearing the cost of negative feelings such that guilt, disappointment or anxiety. This hedonic value of motivated beliefs has been early investigated by Akerlof and Dickens (1982) in a model of cognitive dissonance reduction that relies on two propositions. First, individuals have preferences over beliefs and beliefs thus directly enter as an argument in the utility

function. Second, individuals are able to manipulate their beliefs by selecting information that is the most likely to confirm want they want to believe.

The second value of motivated beliefs is *instrumental* (or *functional*). From a motivational point of view, holding positive beliefs about oneself or one's environment can be a powerful motivator to pursue one's goal, persist in the face of adversity (Bénabou and Tirole, 2006), and even increase one's performance (Compte and Postlewaite, 2004).[4] Holding positive beliefs about one's own value may also help better convince others of it. In that respect, having inflated beliefs about oneself may also be a good instrument to achieve our goals when those later depend on others' decisions (Bénabou and Tirole, 2002). In other words, deceiving oneself helps deceiving others more efficiently (Trivers, 2011). Finally, subjective beliefs may have an instrumental value when they are used as a narrative or an excuse to justify future decisions (Bénabou et al., 2018). For instance, convincing oneself that humans are only partially responsible for climate change prevents from having to change one's habits and acting more responsibly. This instrumental value of motivated beliefs is of particular importance for economists since it leads to questionable behavior that may not only be costly for the individual but for the whole society.

Very few papers have investigated the use of memory as a self-management strategy to sustain these two needs for motivated beliefs. Theoretically, Bénabou and Tirole (2002) have modeled memory manipulation as the equilibrium of a game between the multiple self of a single individual, in which the later can forget information that may damage or threaten his self-confidence. Individuals can,

[4]Compte and Postlewaite (2004) show that emotional valence and performance are positively correlated. Therefore, any positive beliefs that nourish positive emotions may indirectly act as an instrument (i.e. help) to increase performance.

with some probability and possibly at a cost, vary the probability of recalling a given piece of data (Bénabou and Tirole, 2002). Based on Bénabou and Tirole (2002) self-deception framework, Gottlieb (2014) has shown that after observing a negative signal, the decision-maker faces a conflict between forgetting the signal to have a better self-image but make a less appropriate decision, or recalling it and making a better decision. When there is no ex-post decision to make (hedonic value), the self-image factor takes over and the decision-maker recalls a negative (positive, respectively) signal with probability below (above, respectively) the natural percentage predicted by imperfect memory. When there is an ex-post decision (instrumental value) and the self-image and decision-making factors have opposite signs, the amount of memory manipulation depends on the marginal benefit *versus* marginal cost from remembering, which both depend in the utility function on self-image, decision-making and memory costs factors. Very recently, Gödker et al. (2019) have developed a model in which image-concerns form the basis for how information is remembered by the decision-maker. After having observed outcomes of an asset, the investor under-remembers those that are inconsistent with his positive self-image and over-remember the preferred ones. As a consequence, the investor weights asymmetrically each outcome and becomes over-optimistic about the quality of the asset.

Experimental studies confirming or refuting the use of motivated memory as a self-deceptive mechanism are very limited. In 1992, Thompson and Loewenstein (1992) were the first in economics to investigate experimentally selective recalls. They implemented a bargaining game in which subjects in charge of representing different parties in labor negotiation had to reach an agreement with an opponent. If they failed, both parties suffered from a costly strike. The authors found that negotiators showed biased recall of information that favored their own position.

As far as we are aware of, the only other empirical studies on motivated memory in economics are Li (2013, 2017), Chew et al. (2018), Carlson et al. (2018), Zimmermann (2019) and Gödker et al. (2019).

Chew et al. (2018) show that after a delay of several months, individuals exhibit asymmetric recalls of past performance in an IQ test. Individuals forget more their incorrect answers than their correct ones (selective amnesia) but also exhibit false memory encompassing delusion (remembering a positive answer when there was none) and confabulation (transforming a negative answer into a positive one). Using a similar IQ test, Zimmermann (2019) also finds evidence for an asymmetry in the recall of feedback. Precisely, he finds that i) individuals recall negative feedback on their relative performance in the IQ test with less accuracy than positive feedback and ii) individuals who received negative feedback state "I don't recall" more frequently than individuals who received positive feedback. Overall, Zimmermann (2019) shows that individuals manage to suppress feedback that threatens their desire to view themselves as intelligent persons. Li (2017) also tests whether individuals exhibit biased memory in recalling their performance but using a word-entry task instead of an IQ test. Forty days after performing the task, participants are asked to recall their number of mistakes and their performance's rank. The design manipulates whether participants forecast their absolute or relative performance, and whether they receive or not feedback. He finds that both having to forecast performance and receiving feedback eliminate biased recalls. Gödker et al. (2019) also find experimental evidence of a self-serving memory bias, but in financial decisions and not on self-relevant feedback. In their experiment, subjects choose to invest either in a risky asset or a risk-free asset, and observe a series of investment outcomes. Then, either immediately after or one week after depending on the treatment, subjects are asked to recall the observed outcomes. Gödker

et al. (2019) find that subjects who invested in the risky asset under-remember investment losses compared to gains, form overly optimistic beliefs about the future outcomes of the asset and are likely to re-invest in the stock market. In contrast, subjects who invested in the risk-free asset do not exhibit self-serving memory.

As social animals, the demand for positive self-image is not only related to intelligence, performance and/or personal successes but also strongly linked to the desire to appear pro-social, both in one's own eyes (Ariely et al., 2009; Grossman and Van Der Weele, 2017) and in the eyes of others (Bénabou and Tirole, 2002; Battigalli and Dufwenberg, 2007). While some individuals are genuinely other-regarding, others act altruistically not because of their very nature but because they fear to appear selfish to others (DellaVigna et al., 2012). Individuals invest in good deeds to maintain their own view of what kind of person they are, thereby providing evidence that being pro-social plays an important role in self-image building (Bénabou and Tirole, 2006). However, many situations involve a trade-off between favoring others at one's own expense or going first, and people sometimes engage in actions that harm others. This may contradict their demand for pro-social image and even be inconsistent with their own preferences (Banaji and Bhaskar, 2000; Banaji et al., 2004; Chugh et al., 2005; Tenbrunsel et al., 2010). One way to restore these contradictions or inconsistencies between positive self-image and past image-threatening actions is through motivated memory. Li (2013) is the first that studied the existence of memory manipulation in social encounters. He investigated the recollection of decisions in a trust game after various delays but found no evidence that trustees recalled their past decisions asymmetrically depending on whether they reciprocated or betrayed the trustor. Only trustors exhibited asymmetric recalls, betrayed trustors recalling less accurately their decisions than trustors who benefited from reciprocity. Of contrasting

evidence, Carlson et al. (2018) found that misremembering in dictator games was more likely when participants made decisions that fell short of their personal view of fairness.

This thesis investigates experimentally the existence and strength of memory distortion in three economically relevant contexts: social preferences, individual performance and unethical behavior. In each context, it provides a novel body of evidence showing that individuals do use their memory as a self-deceptive strategy to sustain their desire for motivated beliefs. Second, it explores two different determinants underlying memory manipulation: hedonic when memory manipulation only makes oneself look better in one's own eyes, or instrumental when it also helps to justify future decisions. It provides evidence on the existence on memory manipulation for both determinants. Each essay presented here concentrates on the understanding of one determinant of motivated memory in one specific context. More precisely:

- Chapter 1 investigates whether people retrieve their memory self-servingly in social encounters. Do individuals forget the consequences of their actions on others? If so, does it depend on the nature (e.g. selfish or altruistic) of the action? Our results identify a causal effect of the responsibility of pro-social decisions on selective recalls. In contrast, there is no clear evidence of biased memory errors.

- Chapter 2 disentangles between two driving forces that have been proposed as explanations of memory failures for self-relevant information. Why do people exhibit asymmetric recalls for negative and positive feedback on their performance? While the self-enhancing hypothesis claims that people prioritize positive information to enhance their self-image, associative memory states that people just better remember information that is congruent with

their mood. We provide a controlled environment where the two theories predict different outcomes and find that self-enhancing memory takes over mood-congruent memory in the recall of self-relevant feedback.

Chapters 1 and 2 focus on the case where recalls have a purely hedonic value. The decision-maker does not make any ex-post decision, and the only reason for memory manipulation is thus the improvement of his self-view. Chapter 3 explores the case where recalls also have an instrumental value.

- Chapter 3 investigates the relative role of affect and of strategic reasoning in motivated memory, with an application in the domain of unethical behavior. Do individuals exhibit motivated memory of past unethical behavior? If so, is this due to purely hedonic motives and/or strategic reasons? We find that hedonic considerations are not sufficient to trigger memory manipulation. When forgetting serves as a justification for future decisions, however, individuals do motivate their memory.

The following sections present each chapter of this thesis and their original contributions to the existing literature.

Chapter 1: Motivated Memory in Dictator Games

Chapter 1 aims at understanding whether and to what extent individuals manipulate their memory to sustain their demand for pro-social self-image. Like Li (2013), we aim at identifying motivated memory and investigate the selectivity of recalls in social interactions. By contrast to Li (2013), i) we investigate both selective memory (e.g., asymmetric rate of correct recall depending on the desirability of information) *and* biased memory errors (e.g., overly optimistic recalls), and ii) we identify a causal effect of personal responsibility on selective memory.

We designed a laboratory experiment where participants were asked, first, to play a series of binary dictator games and, second, after completing a filler task, to recall the amounts allocated to the receiver. We introduced four treatments in which we manipulated dictators' responsibility for the amount allocated to the receiver (either the dictator chose the amount or the amount was chosen at random by a computer) and the presence of incentives for correct recalls.

Our results show evidence of selective recalls driven by the responsibility of actions. First, when dictators are responsible for the amount allocated to the receivers, their percentage of correct recalls is higher after they chose the altruistic rather than the selfish option. This is not the case when the receiver's amount is selected by the computer. Second, incentivizing correct recalls increases the dictators' percentage of correct recalls but only when they chose the altruistic option and had no effect on accuracy when dictators chose the selfish option. This suggests that people do not completely forget their past decisions but when given a monetary incentive to provide a memory effort, they allocate this effort to retrieve the memory of desirable rather than undesirable information. Finally, when dictators are asked to recall not the receiver's payoff but their own amount, they are also less likely to remember it after choosing the selfish than the altruistic option, showing that selective recalls are not driven by a higher attention paid to the receiver's amount by other-regarding dictators. Together, these results identify a causal effect of the responsibility of pro-social decisions on selective recalls and show that memory errors in social interactions can result from cognitive impairment but also from selective memory.

If our study provides evidence of selective memory, we do not find clear evidence of biased memory errors. Dictators are more likely to over-estimate than under-

estimate the amount allocated to the receiver after choosing the selfish rather than the altruistic option, but the same asymmetry is found when the amount allocated to the receiver is randomly selected by the program. Also, the magnitude of dictators' memory errors is similar regardless of the pro-sociality of decisions and of whether dictators are responsible or not for the amount allocated. Even if a majority of individuals probably prefer to think of themselves as generous rather than egoist, one possible explanation for the absence of biased memory errors is that the dissonance between making selfish decisions when a pro-social alternative is available and maintaining a positive self-image may not be strong enough to generate an internal conflict. Investigating motivated memory in the domain of morality and ethics, where categorical imperative (Kant, 1785) and injunctive norms are more salient, could generate a stronger need for biased memory. Also, in our study individuals could manipulate their memory only for hedonic reasons, i.e., to make themselves look more generous. Such hedonic considerations, alone, may not be sufficient to trigger biased memory errors. Instead, introducing instrumental reasons, i.e., observing situations in which memory manipulation may serve to justify *future* decisions, could help better identify and understand the emergence of biased memory errors and their role in the input-decision process. Exploring these two possibilities is the purpose of the third chapter of this thesis. Chapter 2 continues exploring memory distortions for hedonic reasons, but in the domain of individual performance rather than in social encounters. It investigates the mechanisms behind memory manipulation.

Chapter 2: Mood-driven or Goal-driven Memory Biases?

While several studies find evidence of an asymmetry in recall of self-relevant feedback (Li, 2017; Chew et al., 2018; Zimmermann, 2019), the reasons underlying

this asymmetry are still unclear. In the literature, two main explanations of this recall asymmetry have emerged. On the one hand, motivated memory highlight self-serving explanations, according to which people recall positive information to enhance or protect themselves (self-enhancement effect). On the other hand, theories of associative memory explain asymmetric recall as the result of the enhanced accessibility of positive information and the attenuated accessibility of negative information, when people are in a non-negative mood (mood-congruency effect). Despite the two principles are not mutually exclusive and most existing evidence is consistent with both theories, understanding the driving force of the phenomenon is crucial to predict some suboptimal economic decisions and develop policies aimed at mitigating or removing over-confident behaviors, whether necessary. Indeed, while self-enhancement models grant large meta-cognitive control and thereby suggest the importance of correcting ex-ante beliefs on what is helpful for oneself; mood-congruency models assume an underlying heuristic process without intentionality and thereby suggest that memory biases are simply collateral effect of good affective states.

Chapter 2 is the first attempt to disentangle these two forces which have been proposed as explanations of memory failures for self-relevant information. To identify and disentangle the underlying mechanisms of asymmetric feedback recall, we set a laboratory experiment where both self-enhancing memory and mood-congruent memory offer divergent predictions. Based on the design by Zimmermann (2019), subjects have to perform an IQ test and receive incomplete feedback about their performance relative to their peers. One month later, they come back to the laboratory and are asked to recall their feedback. Before retrieval, we intervene or not on their mood, using Andrade et al. (2015)'s procedure. The laboratory offers a controlled environment where recall accuracy can be carefully

assessed, and mood exogenously manipulated.

Our results provide support for the existence of self-enhancement memory. First, individuals exhibit higher percentage of correct recalls when the feedback was positive than when it was negative. Second, when they do not recall correctly, individuals exhibit positive memory errors. This means that they overestimate the number of positive feedback they received. In other words, individuals exhibit overly optimistic recall of past feedback. Together, these results replicate the findings in Zimmermann (2019). In contrast, we do not find clear evidence of mood-congruent memory, even though our manipulation proves to be effective in inducing the desired affective state. Individuals do not exhibit higher percentage of correct recalls when the feedback to retrieve is congruent to their mood. Overall, these results confirm the effect of self-enhancement memory as a driver of asymmetric recall, but they fail to support any role of mood-congruency.

Our results underline the importance of motivational over affective factors in the formation of optimistic beliefs about the self. The relative dominance of self-enhancement offers direct implications for policies aimed at mitigating or removing biased judgments. Insofar as individuals mostly distort their memory because they consider negative feedback to be potentially harmful, removing ex-ante aversion to negative feedback should be the focus of this agenda.

Chapter 3: Motivated Memory of Unethical Decisions

Chapter 3 investigates the relative role of hedonic motives and of strategic reasoning in motivated memory, with an application in the domain of ethics. Several studies in economics have identified different strategies used by individuals to sus-

tain their value for morality while acting unethically. Individuals avoid knowing the consequences of their behavior (Feiler, 2014; Grossman and Van Der Weele, 2017), exploit norm-uncertainty about lying behavior to justify their own decision to lie (Bicchieri et al., 2019), claim that they have changed (Stanley et al., 2017), shift the blame onto someone else (Bartling and Fischbacher, 2011; Oexl and Grossman, 2013), balance their moral behavior over time (Ploner and Regner, 2013; Gneezy et al., 2014; Cojoc and Stoian, 2014), or use narratives consisting in, for example, downplaying the externalities and/or reinterpret the circumstances of their actions (Bénabou et al., 2018). As far as we are aware of, none of them has explored memory manipulation as another way to sustain humans' value for morality while acting unethically.

Chapter 3 investigates the existence of motivated memory for hedonic and instrumental values in the context of dishonest decision-making. Our study contributes to two strands of the literature. First, it contributes to better understand moral reasoning when facing an opportunity to misbehave. Despite the vast literature on unethical behavior that has flourished in the last decade (for surveys, see Rosenbaum et al., 2014; Irlenbusch and Villeval, 2015; Jacobsen et al., 2018; Abeler et al., 2019), there is no economic study investigating memory manipulation as a self-management mechanism to sustain moral self-image when acting dishonestly. Second, it contributes to the recent but growing economic literature on motivated memory. While existing studies (Li, 2013, 2017; Chew et al., 2018; Zimmermann, 2019; Carlson et al., 2018) investigate motivated memory driven by self-image concerns in the domain of intellectual ability or social preferences, we focus on the memory of dishonest decisions. Moreover, we explore not only whether individuals forget their past unethical behavior to sustain their desire for moral self-image, but also whether they manipulate their memory as an excuse *not*

to engage in subsequent morally responsible behavior. Thereby, we provide the first experimental test of the impact of *anticipated* decisions on memory manipulation. We investigate memory biases not only as a *consequence* of past unethical behavior, but also as an *instrument* to justify future ones.

To study i) whether individuals manipulate the memory of past dishonest choices, and ii) whether they use their memory as an instrument to justify their future decisions, we conducted an on-line experiment where participants first played a repeated mind game allowing them to misreport their outcomes and, three weeks later, were incentivized to recall the distribution of their reports in this game. We varied, across treatments, whether individuals were able to cheat in session one and whether only hedonic or both hedonic and strategic reasons could motivate their memory in the second session.

Our results show that when motivated memory only had an hedonic value –i.e., making oneself look more honest–, dishonest individuals did not recall their past decisions less accurately than participants who were not able to cheat. Although we used a very conservative test, this result suggests that hedonic considerations are not sufficient to trigger a significant memory manipulation in our setting. By contrast, when memory manipulation had an instrumental value –i.e., when individuals were informed, *before* recalling, that they would have a future decision to make on whether or not returning undeserved money–, dishonest individuals recalled their past behavior with less accuracy than when they knew that they would not have any decision to make. This finding suggests that individuals recall selectively as a self-excusing strategy to justify anticipated future decisions. It confirms that memory is involved in the various strategies people use to motivate their beliefs about themselves and justify they can behave immorally while keeping

a positive self-view.

***

The three chapters of this thesis use the experimental method to explore the existence and underlying mechanisms of motivated memory. The benefits from using laboratory experiments to investigate the behavioral determinants of memory distortion are numerous. First, it enables to observe the memory of outcomes induced in the laboratory and thereby allows the experimenter to measure precisely memory errors. This it not necessarily the case in studies investigating self-reported or autobiographical memory. Also, observing both the action and the recollection phases permits not only to identify selective recalls but also to measure the direction and magnitude of memory errors. Second, laboratory experiments enables to exogenously manipulate the nature of the information to be recalled and thereby to test memory retrieval depending on the degree of (un)desirability of information. This is of particular importance when exploring memory as a self-deceptive strategy used by individuals to sustain their desire for motivated beliefs. Finally, it allows the experimenter to control for the experienced duration between the encoding and retrieval phases, and thereby to disentangle the effect of time and the effect of motivation on memory retrieval.

Studying humans' memory by means of laboratory experiments, however, has some specific limitations. First, in laboratory experiments it is common knowledge that the experimenter knows the information that participants are asked to recall. A direct consequence is that participants may face a trade-off between two

self-image serving strategies when asked to recall past behavior. While forgetting undesirable past action or behavior may help sustaining or enhancing a positive self-image, it can also damage it by sending a (negative) signal that one's memory is not performing well. The outcome of such a trade-off may depend on the very nature of the individual. Second, in real life experienced events are rarely recorded and therefore not verifiable. In such uncontrolled environments, memory manipulation may arise more freely than in the lab in which the individual can be confronted with tangible facts and thus experience higher intrinsic costs of memory manipulation. Also, all economic studies that explored memory in the laboratory incentivize correct recalls. While this is crucial to investigate the existence and strength of memory biases, it may provide a conservative measure of memory manipulation. In real life, monetary incentives can sometimes be aligned with forgetting or distorting one's memory. As a provocative example, on can think of an individual who (on purpose but eventually truly) forgets his wallet when going to the restaurant. Less anecdotally, exhibiting memory impairment can lead to alleviated sentences in trial courts (Cima et al., 2002).

Chapter 1

Motivated Memory in Dictator Games[1]

1.1 Introduction

The desire to see oneself in a positive light is a fundamental feature of humans (*e.g.*, Bénabou and Tirole, 2002). People like to think of themselves as good persons. Yet, this demand for positive self-image can be challenged by the fact that most people sometimes behave in ways that they would like to think they did not. The discrepancy between what people do and how they would like to see themselves may create intra-personal conflicts (Conen et al., 1957; Bazerman et al., 1998; O'Connor et al., 2002). One way to restore consistency between positive self-image and past image-threatening actions is through motivated memory. Time gives individuals a wiggle room to forget or distort the memory of actions they would rather not recall (*e.g.*, Moore, 2016). By forgetting or arranging versions of past behavior, motivated memory allows individuals to reconcile the present "want" self with the ex-post "should" self when these two are in conflict (*e.g.*, Bazerman

[1]This chapter is a joint work with Marie Claire Villeval. It has been published in September 2019 in *Games and Economic Behavior*.

et al., 1998; Bénabou and Tirole, 2002).[2] Motivated memory can develop through two channels. *Selective recalls* correspond to asymmetric probabilities of recalling desirable *vs.* undesirable events (Carrillo and Mariotti, 2000; Bénabou and Tirole, 2002; Mullainathan, 2002; Bernheim and Thomadsen, 2005; Gottlieb, 2014; Wilson, 2014) and lead to uncertainty about past self-image threatening actions. *Biased memory errors* refer to the direction and magnitude of errors; they correspond to overly optimistic recalls of past behavior.[3] Motivated memory can thus play in various directions, including selective amnesia but also positive delusion or confabulation. (Chew et al., 2018).

While memory is at the source of any belief formation and updating process,[4] little is known about how individuals use it strategically to sustain their demand for positive self-image, especially in social encounters. Exploring memory biases is important since they may lead to inaccurate statements about oneself, such as overconfidence (*e.g.*, Bénabou and Tirole, 2002), with major implications on the quality of choices. They may also indirectly favor behaviors that are potentially costly for the society: if individuals are able to forget −at least partially− past unethical behavior, they do not have to entirely bear its moral costs. Using a lab-

[2]In psychology, Tenbrunsel et al., 2010 have explored the biased perceptions that people hold of their own ethicality. They argue that the temporal trichotomy of prediction, action and recollection is central to these misperceptions: people predict that they will behave more ethically than they actually do, and when evaluating past (un)ethical behavior, they believe they behaved more ethically than they actually did.

[3]Selective memory corresponds to a different likelihood of recalling or not an event depending on the desirability of this event. Selective memory predicts that the more self-image enhancing is the event, the higher will be the likelihood of recalling it instead of forgetting it. This can be illustrated by the following example. An individual often passes by a homeless person on her way home from work. Sometimes she does not give him money ($0), sometimes she gives a low amount ($1), and sometimes she gives a high amount ($10), equiprobably. Since she likes to think of herself as a generous person, she may better recall the times she gave a positive amount than when she gave nothing. If she underestimates the number of times she gave nothing, she exhibits selective memory. Biased memory is different and refers to the size and the direction of the memory errors. Keeping the same illustration, given inaccurate recalls, if the person is more likely to recall having given $5 when she actually gave $1 than to recall having given $8 when she actually gave $10, memory errors are biased. In other words, selective memory expresses the idea that given that the person recalls the given amounts correctly, the likelihood of recalling the times she gave $0 is lower than the likelihood of recalling the times she gave either $1 or $10. Biased memory errors express the idea that when the person **does not** recall the given amount correctly, the likelihood of overestimating this amount is higher than the likelihood or underestimating it.

[4]In particular, memory manipulations may distort the ability to recall events and thus impair probability assessments (*e.g.*, Hammond et al., 2006).

oratory experiment, our study aims at understanding whether and to what extent individuals manipulate their memory to sustain their demand for pro-social self-image. This relies on two assumptions. First, the demand for positive self-image is linked to the desire to appear pro-social not only in the eyes of others (Bénabou and Tirole, 2006; Battigalli and Dufwenberg, 2007) but also in one's own eyes (Ariely et al., 2009; Grossman and Van Der Weele, 2017). Second, people are able to distort their memory. They can influence the way they encode and recollect information and, if needed, *ex-post* revise their recalls.[5]

Most of the economic literature on this topic is theoretical. Identifying empirically whether individuals use their memory self-servingly is difficult with observational data. Laboratory experiments enable to observe the memory retrieval of outcomes induced in the laboratory. Observing both the action and the recollection phases permits not only to identify selective recalls but also to measure the direction and magnitude of memory errors. In this respect, the rare economic experiments on motivated memory differ from most experiments in psychology that rely on self-reported and/or on autobiographical memory.[6] Moreover, using a controlled environment minimizes the effects of rehearsal and associativeness that strongly impact the individuals' ability to store and recollect information.[7] This permits also to control for the time between the action and the recollection phases, avoiding potential confounds between the effect of time and the effect of motivation on memory retrieval. A last advantage is the control for individual differences

[5] *Memory revisionism* is a process according to which individuals selectively and self-servingly revise the memory of their past behavior to maintain a coherent self-identity (Epstein, 1973; Greenwald, 1980; Markus and Wurf, 1987).

[6] Self-reported or autobiographical memory does not permit to disentangle false memory (when a person recalls something that actually never happened) from motivated memory (when a person experiences a differential percentage of recall or awareness in response to desirable or to undesirable events). In addition, with autobiographical memory the experimenter can hardly check the veracity of the recalled event, which prevents the study of motivated memory at an individual level.

[7] Rehearsal corresponds to the fact that the higher frequency to which an event is remembered makes it easier to remember again. Associativeness corresponds to the fact that the similarity of a past event to a current event makes this latter event easier to recall (Kahana, 2012).

in memory capacity that are hardly observable in natural settings.

To investigate whether individuals use their memory as a self-impression management strategy, we designed an experiment where participants were asked, first, to play a series of binary dictator games and, second, to recall the amounts allocated to the receiver. By introducing social interactions, we differ from the previous economic experiments that mainly investigated how people manipulate their memory about their performance in intelligence tests (Li, 2017; Chew et al., 2018; Zimmermann, 2019), with the exception of Li (2013) who considered trust games and the recent study of (Carlson et al., 2018) on sharing decisions. In our experiment, non selective memory would predict similar percentages of correct recalls and symmetric memory errors for both selfish and altruistic decisions. In contrast, motivated memory predicts that dictators exhibit a different degree of memory accuracy about the amounts given to the receivers, depending on whether they have chosen the option that favors them (the "selfish" option) or the option that favors the receiver (the "altruistic" option). Our intuition is that the choice of the altruistic option leads to a higher memory accuracy and to less biased recalls than the choice of the selfish option because its memory is not self-image threatening. When they have chosen the selfish option, we conjecture that dictators i) exhibit a lower memory accuracy (selective recalls), ii) are more likely to over-estimate and iii) to a larger extent the amount given to the receiver (biased memory errors), compared to when they have chosen the altruistic option. Indeed, dictators who value pro-social self-image may suffer from a higher discrepancy between their self-interested decisions and their desire to see themselves as pro-social when recalling. Memory manipulations may be used self-servingly to reconcile these two selves.

Our contribution to the nascent experimental economic literature on memory is threefold. First, our design allows us to investigate the existence of motivated memory in social interactions in an economic framework. Dictator games engage moral behavior (Konow, 2000; Cappelen et al., 2007), a domain susceptible to motivated memory (Moore, 2016). Moreover, our calibration of the games allows us to identify whether motivated memory is more susceptible to emerge under advantageous or disadvantageous payoff inequality between the dictator and the receiver. Our second contribution is establishing causality between the responsibility of the decisions and motivated memory, by manipulating the dictator's responsibility for the receiver's amount. Contrary to ultimatum or trust games where the responsibility for the final outcome is shared by two players, in dictator games one player bears the entire responsibility for both players' outcomes. This setting does not enable a dilution of responsibility that may substitute to memory manipulations. Our third contribution is estimating selective recalls and biased memory errors (the direction and magnitude of memory manipulation) separately. While most previous experiments (Li, 2013; Chew et al., 2018) offer binary measures −forgetting or recalling−, we can measure the extent to which individuals distort their memory. Also, by manipulating incentives we can, like Zimmermann (2019), disentangle between forgetting and suppression or selective retrieval of past decisions: if past decisions are actually forgotten, incentives should not change the recall accuracy.

In our experiment, participants play 12 binary dictator games. In each game, the dictator has to choose between a selfish and an altruistic option for sharing an amount between himself and a receiver. Across games, we vary both the inequality of payoffs in the two options and whether the dictator or the receiver is in an advantageous position with both options. Then, after performing a distraction task,

players are asked to recall the amounts allocated to the receiver. Participants are not informed of the memory task when playing the dictator games. This design allows us to investigate whether the percentage of correct recalls, the direction, and the magnitude of memory errors differ depending on the option chosen by the dictator.

We introduce four treatments in a between-subjects design. In the Incentive - Receiver's Amount treatment (IRA hereafter), dictators are responsible for the amount allocated to the receiver and correct recalls of the receiver's amount are incentivized. We conjecture that motivated memory increases the dictators' probability to recall after they chose the altruistic rather than the selfish option. The Incentive - Receiver's Amount - Computer treatment (IRAC hereafter) is similar to IRA except that the option is randomly selected by the computer program. Since in this treatment the dictator is not responsible for the amount allocated to the receiver, we conjecture no difference in recalls between selfish and altruistic options. The No-Incentive - Receiver's Amount treatment (NIRA hereafter) is similar to the IRA treatment, except that correct recalls are not incentivized. If individuals forget past decisions, introducing or removing incentives should not affect the accuracy of recalls; in contrast, if accuracy depends on incentives, it suggests that people either selectively suppress the past decisions they are not so proud of, or make a greater effort to retrieve past image-enhancing decisions. Finally, if selfish dictators make more memory errors, it might be because of motivated memory or because they paid less attention to the receiver's amount when making decisions. We ran an Incentive - Dictator's Amount treatment (IDA hereafter) that is similar to the IRA treatment, except that participants have to recall the amount allocated to the dictator. We conjecture that biased dictators should exhibit a different percentage of correct recalls and a different magnitude of memory errors depending

on their chosen option not only when they have to recall the receiver's amount, but also when they have to recall their own amount.

Our results show evidence of selectively accurate *vs.* inaccurate recalls driven by the responsibility of actions. First, when dictators are responsible for the amount allocated to the receivers, their percentage of correct recalls is higher after they chose the altruistic rather than the selfish option. This is not the case when the receiver's amount is selected randomly. Second, incentivizing correct recalls increases the dictators' percentage of correct recalls when they chose the altruistic option but not when they chose the selfish option. This suggests that people do not forget their past decisions but when given a monetary incentive to provide a memory effort, they allocate this effort to retrieve the memory of desirable rather than undesirable information. Finally, in the IDA treatment dictators are also less likely to remember their own amount after choosing the selfish than the altruistic option. In contrast, we do not find clear evidence of biased memory errors. Dictators are more likely to over-estimate than under-estimate the amount allocated to the receiver after choosing the selfish rather than the altruistic option. However, the same asymmetry is found when the amount allocated to the receiver is randomly selected by the program. Also, the magnitude of dictators' memory errors is similar regardless of the pro-sociality of decisions and of whether dictators are responsible or not for the amount allocated.

Overall, these findings identify a causal effect of the responsibility of pro-social decisions on selective recalls but not on biased memory errors. Individuals have a less accurate memory of past behavior when they have been selfish but they do not exhibit overly optimistic recalls of their past behavior. These selective recalls in social interactions are consistent with theoretical and empirical studies establish-

ing an asymmetric recall of feedback depending on whether people receive good news or bad news about their relative performance (Bénabou and Tirole, 2002; Li, 2017; Chew et al., 2018; Zimmermann, 2019). These findings show that memory errors can result from cognitive impairment but also from motivated biases.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the experimental design and procedures. Section 4 outlines the behavioral conjectures. Section 5 reports the results and section 6 provides robustness tests. Section 7 discusses these findings and concludes.

1.2 Related Literature

Psychologists have intensively investigated the individuals' tendency to selectively forget self-threatening information. They have shown that people are more likely to recall their successes than their failures (Korner, 1950; Mischel et al., 1976), they have self-serving recollections of their past performance (Crary, 1966), they exhibit poorer recall of negative *vs.* positive self-relevant information (Green and Sedikides, 2004; Sedikides and Green, 2009), and they recall more accurately favorable than unfavorable feedback (Story, 1998). In the context of social interactions, people sometimes engage in actions that harm others, which contradicts their demand for pro-social image and may even be inconsistent with their own preferences (Banaji and Bhaskar, 2000; Banaji et al., 2004; Chugh et al., 2005; Tenbrunsel et al., 2010). Since people are threatened by information that has undesirable implications for their self-image, poor recall of this information may help think of past behavior under a positive light (Moore, 2016). For example, Stanley et al. (2017) have shown that recalled actions that involve emotional harm are per-

ceived as more morally wrong when participants are put in the shoes of the actor than when put in the shoes of an observer. Also, people have less clear memory of their own unethical experiences than of their ethical experiences, while they recall others' ethical and unethical actions similarly (Kouchaki and Gino, 2016).

While the economic literature modelling cognitive limitations in recalls and their impact on belief formation and decision-making is substantial (Dow, 1991; Piccione and Rubinstein, 1997; Mullainathan, 2002; Bénabou and Tirole, 2004; Brunnermeier and Parker, 2005; Bénabou and Tirole, 2006; Wilson, 2014; Bordalo et al., 2017), very few papers have investigated the use of memory as a self-deceptive mechanism. In a model where individuals can vary the probability of recalling a given piece of data, Bénabou and Tirole (2002) show that individuals have an incentive to forget signals that undermine long-term goals (for motivational reasons) or lower self-esteem (for affective reasons). In a multiple-self model, Gottlieb (2014) shows that after observing a negative signal, the decision-maker faces a conflict between forgetting the signal and having a better self-image, or recalling it and making a better decision. When there is no ex-post decision to make, the self-image factor takes over and the decision-maker recalls a negative (positive, respectively) signal with probability below (above, respectively) the actual percentage. Our study takes root in these models, focusing on the case where signals have a purely hedonic or affective value. The decision-maker does not make any *ex-post* decision and the only reason for memory manipulation is the improvement of his self-view.

Economists recognize the role played by memory in the maintenance of self-image in theoretical models, but they have provided limited empirical evidence. As far as we are aware of, the only empirical studies on motivated memory in

economics are Li (2013); Dessi et al. (2016); Li (2017); Chew et al. (2018); Zimmermann (2019) and Carlson et al. (2018). Chew et al. (2018) show that after a delay of several months, individuals exhibit asymmetric recalls of past performance in an IQ test. They forget more their incorrect answers than their correct ones. However, before having to recall whether their answer was correct or not they were shown the correct answer. Thus, they may distort their recalls but also deceive themselves to self-signal a higher ability without using their memory, especially since the time between the action and the recollection was from months to a year (see, *e.g.*, Mijović-Prelec and Prelec (2010) for a model of self-deception as self-signalling). In our experiment people do not receive any feedback between the decision and the recollection phases that both take place within the same session. Thus, they have a higher chance to recall the amounts given to the receivers, which should limit direct self-signalling deception. Also, we can explore the magnitude of memory errors.

Zimmermann (2019) investigates the underlying mechanism of motivated beliefs and provides evidence of asymmetry in the recall of feedback on past relative performance in an IQ test. Different treatments manipulate the incentives for correct recalls and the time between feedback and the second elicitation of beliefs about one's rank in a group. People adjust their posterior beliefs just after receiving feedback on their performance, but when these beliefs are elicited one month later rather than immediately, people who received positive feedback keep high beliefs whereas those who received negative feedback return to their prior beliefs. By varying incentives, Zimmermann (2019) find that people manage to suppress feedback that threatens their desire to view themselves as intelligent persons. Li (2017) also tests whether individuals exhibit biased memory in recalling their performance but using a word-entry task instead of an IQ test. Forty days

after performing the task, participants are asked to recall their number of mistakes and their performance rank. The design manipulates whether they forecast their absolute or relative performances, and whether they receive or not feedback. Both having to forecast performance and receiving feedback eliminate biased recalls.

Like these studies, we aim at identifying motivated memory and we investigate the selectivity of recalls. We also manipulate the existence of incentives for accurate recalls to test for forgetting or suppression of past decisions. In contrast to these studies, we manipulate exogenously the dictators' responsibility of decisions and therefore we are able to identify a causal effect of decisions on selective memory. We also focus on the memory of other-impacting decisions and we explore another side of individuals' desired self-view: the demand for pro-social self-image. Motivated memory on pro-sociality has been almost unexplored. An exception is Li (2013) who investigates the recollection of decisions in a trust game after various delays.[8] Betrayed trustors have a lower recall accuracy, while those who benefit from kind acts remember perfectly. In contrast, the probability of trustees to recall their past decisions is the same, regardless of whether they reciprocated or betrayed the trustor. We differ from this study in several respects. We use dictator games instead of trust games because the dictator bears the full responsibility of the receiver's payoff, which we assume has a key role in triggering selective memory and allows us to identify a causal effect on selective memory. Also, we do not manipulate the time between decisions and recalls but we explore both selective recalls and biased memory errors, which highlights the underlying mechanism of motivated memory. A key point is indeed to investigate not only whether participants recall or not, but also whether recalls are systematically biased self-servingly in one direction and whether the magnitude of the bias depends

[8]Dessi et al. (2016) study the ability to recall information about friendship networks, but not in the perspective of exploring memory as a self-view management mode.

on the pro-sociality of the decision. Finally, we have been aware of a recent study by Carlson et al. (2018) that, like us, investigate motivated memory in dictator games. The two studies developed independently. The authors found that misremembering in dictator games is more likely when participants made decisions that fall short of their personal view of fairness. By contrast, we focus on the role of personal responsibility in decision-making on motivated memory and we compare whether memory errors differ when participants have to recall the amount given to another person and the amount kept for themselves.

1.3 Experimental Design and Procedures

We describe the design of the experiment before detailing the procedures.

1.3.1 Design

Our experiment consists in four parts. In part 1, participants play dictator games. In part 2, they perform a distraction task used to wipe out the instant memory of part 1. In part 3, in most treatments they are asked to recall the amount allocated to the receiver in each game played in part 1. In part 4 we measure the participants' general memory capacity. Instructions are included in Appendix 1. We now describe each part in detail.

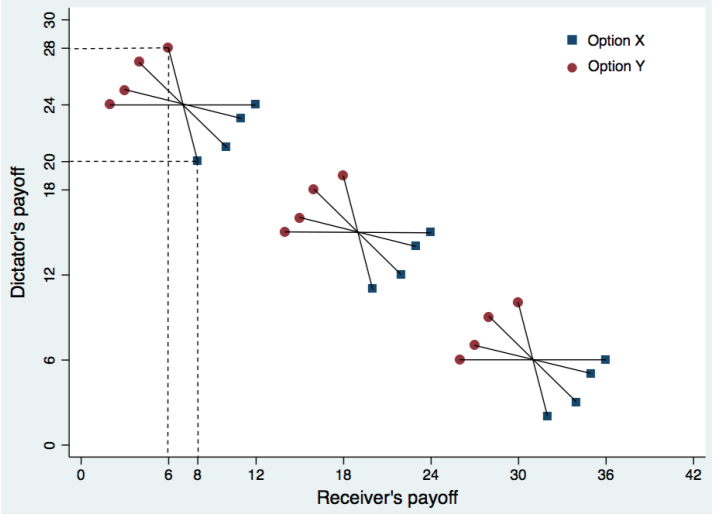
Part 1: Dictator Games

In part 1, participants play twelve binary dictator games, as described in Table 1.1. Half of the participants are dictators (players A), the other half are receivers (players B). Roles are randomly assigned at the beginning of the part and kept

constant for the twelve games.[9] Dictators and receivers are randomly re-matched after each game. In each game, the dictator has to choose one of two options. Option X pays X_a to the dictator and X_b to the receiver. Option Y pays Y_a to the dictator and Y_b to the receiver. The receiver is passive. At the end of the session, one game is randomly selected for payment. Participants are not informed that they will be asked to recall the receiver's amounts in part 3. To avoid any possible confound, each game is unique and each receiver's amount (X_b or Y_b) appears only once. Figure 1.1 illustrates the games. The calibration is inspired by Bruhin et al. (2018). Each star represents a set of four dictator games in different payoff spaces. In the top-left payoff space, dictators are always in an advantageous position: their amount is always higher than the receiver's amount, regardless of the chosen option. In the middle payoff space, the position depends on the chosen option. In option X dictators are in a disadvantageous position while they switch to an advantageous position in option Y. In the bottom-right payoff space, dictators are always in a disadvantageous position, regardless of the chosen option. Hereafter, option X is called the "altruistic" option and option Y the "selfish" option. In the altruistic option, the dictator's amount is always lower than in the selfish option.

A crucial aspect of the design is that participants must pay sufficient attention to the games to encode and to be able to recall the amounts in part 3. To that aim, we implemented some rules. First, the screens that display the two options are frozen during five seconds before dictators can enter their decision. Second, dictators have to type in the dictator's and the receiver's amounts in the chosen option. Then, the option chosen by the dictator remains visible on the receiver's

[9] We decided not to play under the veil of ignorance for two reasons. First, deciding under uncertainty about one's role could have affected the measurement of other-regarding preferences (Casari and Cason, 2009; Iriberri and Rey-Biel, 2011). Second, choices under role uncertainty are less image threatening both before and after role assignment. Before, because the player does not know whether his decision is going to be implemented and he may thus distantiate himself from the responsibility of outcomes. After, because once roles have been assigned, the dictator can persuade himself that the others have made the same selfish choices, which may reduce guilt and the need to bias memory.



Notes: Each game is represented by a line that connects options X and Y. The slope of the line represents the cost for the dictator of increasing the receiver's amount. Each of the three stars represents a set of four games in different payoff spaces. In the top-left space, dictators are always in an advantageous position. In the middle space, the position depends on the chosen option. In the bottom-right space, dictators are always in a disadvantageous position. *Example (dashed lines)*: option X yields 20 ECU to the dictator and 8 ECU to the receiver.

Figure 1.1: The Dictator Games

screen for five seconds. For symmetry receivers have also to type in the same amounts. Typing the amounts increases the probability to recall these amounts, as writing down a statement helps memorize it (see, *e.g.*, Naka and Naoi (1995) and Skinner et al. (1997)).

Part 2: Filler Task

Part 2 introduces a filler task (solving mazes during eight minutes - see Appendix 1) that requires attention and concentration and which purpose is to distract participants from the previous task and allow some forgetting. Drawing the participants' attention away from the previous dictator decisions may open a wiggle room for memory manipulation. Each maze solved pays €0.25.

Table 1.1: The Binary Dictator Games

Games	Option X Altruistic	Option Y Selfish	Relative position of the dictator
1	(2, 32)	(10, 30)	Disadvantageous
2	(3, 34)	(9, 28)	Disadvantageous
3	(5, 35)	(7, 27)	Disadvantageous
4	(6, 36)	(6, 26)	Disadvantageous
5	(11, 20)	(19, 18)	Mixed
6	(12, 22)	(18, 16)	Mixed
7	(14, 23)	(16, 15)	Mixed
8	(15, 24)	(15, 14)	Mixed
9	(20, 8)	(28, 6)	Advantageous
10	(21, 10)	(27, 4)	Advantageous
11	(23, 11)	(25, 3)	Advantageous
12	(24, 12)	(24, 2)	Advantageous

Notes: The first numbers in parentheses display the dictator's amounts, the second numbers the receiver's amounts. The receiver's amount is always higher with option X. The dictator's amount is always higher (or equal) with option Y.

Part 3: Memory Task

Part 3 introduces the memory task. For each dictator game played in part 1, participants, regardless of their role, are asked to recall and report the amount allocated to the receiver in the selected option.[10] For each game the screen displays the two options, but for the option actually chosen by the dictator in part 1, the receiver's amount is replaced by a question mark. All the amounts to be recalled are between 2 and 36. However, to give each amount a chance to be over- and under-estimated, we allowed the recalls to lie in the interval 0 to 38, inclusive. Participants are informed that the amounts to recall are within this range. This task allows us to measure both selective recalls and biased memory errors.[11]

[10] Participants were not informed about the memory task before part 3 because it might have impacted not only their recall accuracy but also their choice of option. Indeed, if they anticipate negative utility due to self-image threatening decisions, they may make less selfish choices strategically and thus, they have no incentives to bias their recalls. In addition, knowing that they will be paid for correct recalls, they could act strategically by choosing not the option they prefer but the option whose amounts are easier to recall.

[11] Note that over-estimating the amount given to the receiver, if any, could be driven by motivated memory but also possibly by social image concerns *vis-a-vis* the experimenter. Eliciting recalls a month later instead of within the same session would help investigate whether memory selectively fades over time; however, it could also reinforce social image concerns. When recalls are elicited in the same session, incentives might still be salient and the trade-off between a better social image and a higher payoff might be pronounced. In contrast, a month later incentives have been received, probably spent, and might thus appear less salient; thus, the relative importance of social image might increase over time.

Games are displayed in a random order independent from their order in part 1. Two recalls are selected randomly. Each correct recall pays two Euros, a correct recall plus or minus one unit pays one Euro, and otherwise participants neither earn nor lose anything.

Dictators are asked to recall the amount allocated to the receiver in the selected option and not the chosen option for three reasons. First, the time span between the decision and the recollection may be too short to observe forgetting. In Li (2013), less than 5% of the players forgot their choice when the decision and the recollection were on the same day. Having to recall the amount left to the other player is harder and leaves room for forgetting. Second, if a participant does not recall his decision, he may simply play the game again. If preferences are stable over time, he should be able to find the option he had chosen without recruiting any memory effort. Third, asking the receiver's amount allows us to measure both the direction and the magnitude of memory errors, if any, and not only the existence of selective recalls.[12]

Part 4: Elicitation of Memory Capacity

The capacity to memorize may be heterogeneous across individuals. Thus, in the last part we elicit participants' memory capacity in an individual environment. To avoid any confound with the memory task in part 3, the new task does not involve numbers but tests verbal memory. It is adapted from one of the three paradigms used to study memory performance (Bordalo et al., 2017): the free recall test (see *e.g.*, Murdock Jr (1962); Tulving et al. (1972)). This part is made of three rounds.

[12]Our memory task is cognitively demanding. We could have used instead a standard one-shot dictator game and increased the time span between decisions and recalls. However, in Li (2013) even after 43 days, more that 85% of the participants recalled their choice in a trust game. Using a repeated game increases the space for forgetting. Moreover, at the time of recollection, participants could have played the game another time instead of trying to remember their decision. Thus, any difference between recalls and decisions could be attributed to motivated memory but also to a variation of preferences over time.

In each round, participants have to read and memorize a sequence of 15 random words. Each word is displayed one by one on the screen for two seconds. Then, participants are asked to recall as many words as possible. They receive no feedback on their performance until the end of the session. They are paid according to their performance in a round selected at the end of the session. Each correct recall pays €0.25. Finally, participants have to fill out a standard demographic questionnaire.

We acknowledge that this measure is imperfect since it tests verbal memory whereas our main task is about memorizing numbers, and it was administered at the end of the experiment when subjects were possibly tired. But administering the test at the beginning of the session could have primed the subjects about the nature of the main task. Despite its limitations, this measure remains informative since psychologists have shown that verbal empan (the highest number of words that an individual is able to recall) and digit empan (the highest number of digits that an individual is able to recall) are significantly correlated within individuals (Hilton, 2006).

Treatments

Our four between-subjects treatments are summarized in Table 1.2. The Incentive - Receiver's Amount treatment (IRA) is the baseline. Dictators choose the amount allocated to the receiver, have to recall this amount, and are paid for accurate recalls. The Incentive - Receiver's Amount - Computer treatment (IRAC) is similar to IRA, except that the option in the dictator games is always selected randomly by the computer program instead of being chosen by the dictator. Dictators in this treatment bear no responsibility for the receiver's outcome. The comparison between the IRA and IRAC treatments indicates whether the responsibility for

decisions triggers motivated memory, if any.

The No-Incentive - Receiver's Amount treatment (NIRA) is similar to IRA, except that recalls are not incentivized. The comparison between the NIRA and IRA treatments allows us to test whether individuals erase definitely some decisions from their memory, or whether they either suppress or retrieve them selectively. We expect that recalls are less selective and biased when manipulation is costly. If individuals actually forget, incentives should not affect the accuracy of recalls regardless of the option (see Zimmermann, 2019). If selfish choices are recalled less when incentives are absent rather than present, this indicates that individuals suppressed them; if altruistic choices are recalled more when incentives are present than when they are absent, this indicates that individuals allocate their memory effort selectively.

Finally, in the Incentive - Dictator's Amount treatment (IDA) participants have to recall the amount allocated to the dictator instead of the amount allocated to the receiver. Recalls are incentivized like in IRA and IRAC. This treatment should control for the fact that social preferences may condition the attention paid to the receiver's amount, and thus the memory of it. If any difference in memory accuracy across the chosen options in IRA is driven by differential attention, the percentage of correct recalls should not differ between the selfish and the altruistic options when dictators have to recall their own amount. If memory accuracy depends on self-image concerns, the difference in accuracy between the recalls of selfish *vs.* altruistic decisions should be similar in this treatment and in IRA.

Table 1.2: Summary of Treatments

Treatment	IRA	IRAC	NIRA	IDA
Active dictator	Yes	No	Yes	Yes
Incentives for accurate recalls	Yes	Yes	No	Yes
Recall of the receiver's amount	Yes	Yes	Yes	No

Notes: IRA: Incentives - Receiver's Amount. IRAC: Incentives - Receiver's Amount - Computer. NIRA: No-Incentives - Receiver's Amount. IDA: Incentives - Dictator's Amount.

1.3.2 Procedures

The experiment was programmed using Java language. It was conducted at GATE-Lab, Lyon, France. A total of 620 participants were recruited from our subject-pool, mainly from local engineering and business schools, using hroot (Bock et al., 2014). 158 participated in the IRA treatment, 154 in the IRAC treatment, 146 in the NIRA treatment and 162 in the IDA treatment. Table A2.2 in Appendix 2 summarizes the participants' characteristics in each treatment.

Upon arrival, each participant was randomly allocated to a terminal. Instructions for each part were self-contained and displayed on the participants' screen at the end of the previous part. No feedback on performance or earnings was provided until after all parts were completed. The use of paper, pen or mobile phone was prohibited. Sessions lasted on average 55 mins. At the end of the session, participants were paid individually in cash in a separate room. They earned on average 15.01 Euros (S.D. 2.79), including a 5-Euro show-up fee.

1.4 Behavioral Conjectures

The following section formulates four behavioral conjectures regarding the asymmetry of dictators' recalls conditional on the selected option (selective recalls), the

direction and the magnitude of memory errors (biased memory errors).

At the time of the decision, dictators may prefer the option that maximizes their own payoff. But at the time of the recollection, they may prefer to recall that the chosen option was more generous to the receiver than it was actually.[13] When dictators have chosen the altruistic option, recalling correctly how much they gave to the receiver has no undesirable implications in terms of self-image. In contrast, when dictators have chosen the selfish option, recalling accurately the amount given to the receiver may conflict with the desire to see themselves as pro-social. In this case, dictators have some motivation to exhibit poorer recall of the amount actually allocated to the receiver. This is consistent with Benabou and Tirole (2002) where individuals are motivated to forget signals that undermine their long-term goals (motivational reason) or lower their self-esteem (affective reason). Here, motivated memory can only respond to affective reasons when the individual sends a signal to himself about his nature when he chose his options in the dictator games.

In contrast to IRA, in IRAC the receiver's amount is selected randomly by the program. Since the dictator is not responsible for the amount allocated to the receiver, the selection of the selfish option is not self-image threatening. We conjecture that the responsibility for the receiver's amount in IRA leads to selective recalls, *i.e.* a difference in the probability of an accurate recall after the choice of the selfish option *vs.* the altruistic option. In contrast, we do not expect any difference in this probability when the option has been selected by the program in IRAC.

[13] Tenbrunsel et al. (2010) use the "want/should" framework to explain the bounded ethicality that arises from temporal inconsistencies. They posit that the "should" self, −characterized by intentions and beliefs on how one ought to behave−, dominates during the prediction and recollection phases, whereas the "want" self, −characterized by a relative disregard for ethical considerations−, dominates during the action phase.

We state our first conjecture as follows:

Conjecture 1 (Selective Recalls) *Because people prefer receiving good rather than bad signals about their own nature, the percentage of accurate recalls is higher when the dictators chose the altruistic option than when they chose the selfish option, whereas these percentages are the same when dictators bear no responsibility in the choice of option.*

If individuals actually forget their past decisions, the accuracy of recalls should not vary according to the presence (IRA) or absence (NIRA) of monetary incentives. If they do not forget, monetary incentives are expected to increase the accuracy of recalls. There may be two effects. First, incentives may increase the individuals' effort to retrieve the memory of decisions that give them a positive self-image more than the memory of those that threaten their self-image. As a result, introducing incentives should increase the accuracy of recalls when the altruistic option has been chosen. Second, incentives may discourage people from suppressing the memory of decisions they are not so proud of because of the opportunity cost for not being accurate. Indeed, as modeled in Bénabou and Tirole (2002), incentives introduce a trade-off between the affective benefits from biasing beliefs in a self-serving way and the monetary incentives for accurate beliefs: on the one hand, a correct recall increases payoff but may threaten self-image, on the other hand, forgetting satisfies the demand for positive self-image but leads to give up the bonus for correct recalls. As a result, introducing incentives may increase the accuracy of recalls when the selfish option has been chosen.

This leads to our second conjecture about selective recalls:

Conjecture 2 (Incentives and Selective Recalls) *The percentage of dictators' correct recalls is higher when correct recalls are incentivized than when they are not incentivized, regardless of the option.*

The following two conjectures are related to biased memory errors. Psychologists have shown that individuals not only tend to forget self-image threatening information but also sometimes arrange past events or even create false memories (Gonsalves and Paller, 2002; Gonsalves et al., 2004; Chrobak and Zaragoza, 2008). Biasing one's memory self-servingly allows individuals to reconcile their actual action with the action they, ex-post, would have preferred to think they made. Our design allows us to investigate not only whether participants recall correctly the amounts allocated to the receivers, but also the direction and the magnitude of memory errors. When participants do not recall the exact amount, they can either over-estimate or under-estimate it, and to a greater or a lesser extent. The difference between the recalled and the actual amounts across decisions allows us to disentangle simple errors from biased memory errors. Simple errors should give similar percentages of over- and under-estimation and a similar magnitude of memory errors across options and across treatments. In contrast, if dictators manipulate their memory to appear pro-social to themselves, they are expected to over-estimate more often the receiver's amount when they chose the selfish option than when they chose the altruistic option, and to a larger extent.

This leads to our third and fourth conjectures:

Conjecture 3 (Direction of Memory Errors) *The percentage of over-estimated recalls is higher when dictators chose the selfish option than when they chose the altruistic option, while no difference is expected when dictators bear no responsibility in the choice of option.*

Conjecture 4 (Magnitude of Memory Errors) *Dictators' recalls over-estimate the amount given to the receiver to a larger extent when they preferred the selfish option to the altruistic option, while no difference is expected when dictators bear no responsibility in the choice of option.*

1.5 Results

We present four results that correspond to the four conjectures. The first result analyzes the impact of dictators' responsibility on selective recalls. The second result presents the impact of monetary incentives on selective recalls. Results three and four investigate biased memory errors by exploring the direction and the magnitude of these errors, respectively. In this analysis, a recall is defined as correct if the recalled amount is equal to the actual amount plus or minus one unit.[14],[15] When a recall is incorrect, a memory error is defined as the difference between the recalled amount and the amount actually transferred by the dictator.

We introduce our first result:

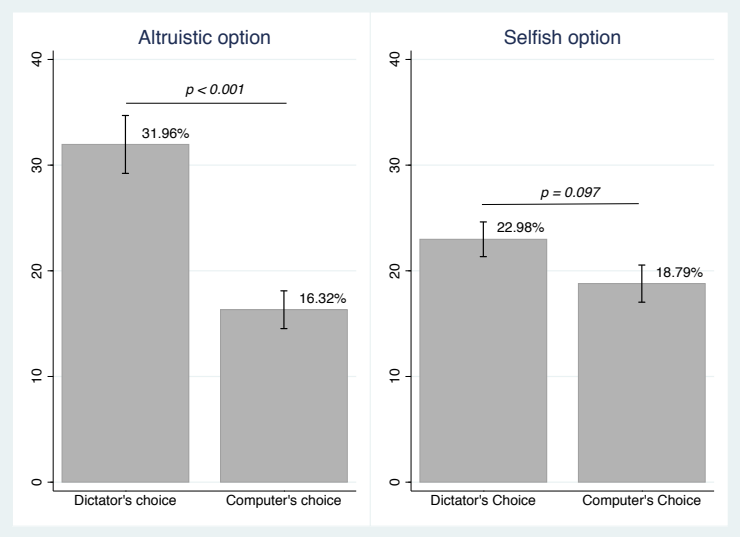
Result 1 (Selective Recalls) *The percentage of dictators' correct recalls is higher when they chose the altruistic option than when they chose the selfish option; this is not the case when dictators bear no responsibility in the choice of option.*

To support Result 1, we provide three types of analyses. Two of the three analyses support Conjecture 1.

Support for Result 1: We start with the most conservative non-parametric tests. Figure 1.2 displays the average percentage of dictator's correct recalls in

[14]We replicated our analysis using both a stricter definition (a recall is defined as correct if it matches exactly the actual amount) and a less strict definition (it is defined as correct if it deviates from the actual amount by up to two units). All specifications qualitatively confirm the main results (see Appendix 4).

[15]We restrict our analysis to the dictators' recalls although we also elicited the receivers' recalls of dictators' choices. Comparing dictators' and receivers' recalls cannot provide a clean identification of dictators' motivated memory because receivers may also motivate their memory, albeit for different reasons. For example, they may remember better the altruistic decisions because they may want to believe that they are surrounded by altruistic people, because they made them happier, or because they try to experience the positive anticipated utility from expected higher future payoffs. Interested readers can find an analysis of the receivers' recalls in Appendix 5.



Notes: The Figure displays the percentages of dictators' correct recalls depending on the option (altruistic or selfish) chosen by the dictator in IRA or by the program in IRAC. p-values are from Mann-Whitney tests.

Figure 1.2: Average Percentage of Dictator's Correct Recalls in IRA and IRAC

the IRA and IRAC treatments, by option, and Table 1.3 summarizes descriptive statistics on recalls (the raw individual decisions and recalls are displayed in Figure A3.5 in Appendix 3).[16] In IRA, dictators recall accurately the amount allocated to the receiver 31.96% of the time when they have chosen the altruistic option and 22.98% of the time when they have chosen the selfish option. These percentages go in the direction of Conjecture 1; however, a Wilcoxon signed-rank test (W test, hereafter) with one observation per subject per type of decision, shows that the difference is not significant (p=0.517).[17] In the IRAC treatment in which the option is selected by the program, dictators recall accurately the amount allocated to the receiver 16.31% of the time when the altruistic option has been selected, and 18.79% of the time when the selfish option has been selected. This difference

[16]Table A2.1 in Appendix 2 presents the relative frequency of the selfish choice in each game, by treatment. This frequency is 69.30% in IRA, 53.57% in IRAC, 69.75% in NIRA, and 67.70% in IDA and it varies across games, which gives opportunities for memory manipulation. Table A2.3 in Appendix 2 summarizes statistics on behavior in the four parts of the experiment and in the final questionnaire, by treatment.

[17]In all non-parametric tests reported in this paper, the average recall of each individual gives one independent observation, and all tests are two-sided.

is marginally significant (p=0.088) and goes in the opposite direction of what is observed in IRA. Comparing IRA and IRAC reveals that the percentage of correct recalls for the altruistic option is significantly higher in IRA (Mann-Whitney tests −M-W hereafter−, p<0.001); this is also the case for the selfish option but to a much lower extent (p=0.097).

Table 1.3: Summary Statistics - Dictators' Recalls

	IRA (1)	IRAC (2)	NIRA (3)	IDA (4)
Percentage of correct recalls, by chosen option				
Alt. option	31.96% (291)	16.31%***(429)	25.28%** (265)	42.36%* (314)
Self. option	22.98% (657)	18.79%* (495)	24.06% (611)	28.27% (658)
p-values	*0.517*	*0.088*	*0.320*	*0.010*
Percentage of over-estimated recalls, by chosen option				
Alt. option	31.82% (198)	23.96% (359)	30.30% (198)	52.49%*** (181)
Self. option	54.74% (506)	62.44%*** (402)	54.96% (464)	38.98%*** (472)
p-values	*<0.001*	*<0.001*	*<0.001*	*<0.001*
Magnitude of absolute memory errors, by chosen option				
Alt. option	5.06 (291)	7.27*** (429)	6.08 (265)	3.10*** (314)
Self. option	5.75 (657)	7.02*** (495)	5.58 (611)	4.11*** (658)
p-values	*0.665*	*0.566*	*0.355*	*0.035*

Notes: In the non-parametric tests, each individual gives one independent observation. Numbers in parentheses indicate the number of individual observations. The p-values in lines compare recalls when the altruistic *vs.* selfish options have been chosen, using W tests. The stars in columns come from pairwise treatment comparisons with IRA taken as the reference category, using M-W tests. *p<0.10, **p<0.05, ***p<0.01.

Our second test of Conjecture 1 examines whether the accuracy of recalls varies across various types of dictators. In each treatment, we split the sample of dictators based on the median frequency of selfish choices.[18],[19] In IRA, the more selfish dictators (those who chose the selfish option in more than eight games, N=41) exhibit a lower average percentage of correct recalls (21.34%) than the less

[18]In IRA, the median frequency of selfish choices is 8. It is 6 in IRAC since the options were selected at random.

[19]The design of the 12 games allows us to identify more than two types of dictators, *i.e.*, spiteful, altruistic, inequality averse individuals and social welfare maximizers. However, we did not use these categories in our analysis of motivated memory for several reasons. In particular, there are no obvious mechanisms that would justify conjectures on dictators' recalls depending on their type. For example, between social-welfare maximizers and inequality-averse individuals, it is not clear who should be more susceptible to exhibit motivated memory. Moreover, such a classification requires that players exhibit a consistent pattern of decisions across the 12 games, which is not systematically observed.

selfish dictators (N=38; 30.48%). The difference is highly significant (M-W test, p=0.006). This is not the case when dictators are not responsible for the amounts allocated to the receivers. Indeed, in IRAC the percentage of correct recalls is 17.34% for dictators with a number of selfish choices above or equal to six (N=37) and 17.92% for those with a number of selfish choices below six (N=40). These percentages are not significantly different (M-W, p=0.810). This analysis supports Result 1: dictators exhibit selective recalls when they are responsible for the amount given to the receivers. The difference in recall accuracy between the more selfish and the less selfish active dictators cannot be explained by differences in memory capacity. Indeed, more selfish dictators (N=41) do not differ from less selfish ones (N=38) in terms of memory capacity in the verbal memory task. On average more selfish dictators remember 24.66 words correctly out of 45 and less selfish dictators remember 25.29 words correctly (M-W test, p=0.640).

Our third test of Conjecture 1 is based on a regression analysis that controls for the characteristics of the games and of the individuals. Table 1.4 reports the marginal effects from Logit regressions in which the dependent variable is equal to one if the recall is correct and zero otherwise. Robust standard errors are clustered at the individual level. In model (1) the independent variables include the four treatments (with IRA as the reference category) and the option chosen by the dictator (selfish *vs.* altruistic) in order to test the presence of selective recalls. They also include the three sets of games indicating whether the dictator was in an advantageous or a disadvantageous position regardless of his choice, or in a mixed situation depending on his choice (with the advantageous category taken as the reference). This allows us to test whether the demand for motivated memory is lower when the dictator is always in a disadvantageous position in a game because it might be easier to justify a selfish choice in this setting. The independent

variables also include the time spent to enter the recall and the game orders in part 1 (dictator games) and in part 3 (recalls) because they may impact memory accuracy, as attention may have decreased over time. Finally, we control for the performance of the participant in the verbal memory task performed in part 4 and for various demographic variables (age, male and educational attainment, as measured by the number of years of study after high school). Models (2) to (5) replicate model (1) for each treatment separately.

Table 1.4 supports Result 1 on the existence of selective recalls. Having to recall the amount given to the receiver when the selfish option has been chosen in IRA decreases significantly (at the 5% level) the likelihood of a correct recall (model (2)).[20] This is not the case in IRAC: there is no difference in the likelihood of recalling accurately when the program has selected the selfish or the altruistic option (model (3)). The relative position of the dictators in the game does not affect memory, as the various sets of games have no significant effect. This confirms descriptive statistics: the average percentage of correct recalls is 26.61% when dictators are in a disadvantageous position, 25.24% when they are in a mixed position, and 23.95% when they are in a disadvantageous position, with no significant differences in pairwise comparisons (W tests, $p>0.010$). The conclusion remains if the analysis is restricted to the cases in which dictators select the selfish option: when in an advantageous position, they do not exhibit less memory accuracy than in any other position (see Table A2.4 in Appendix 2). This may result from the fact that the games were presented in random order, which probably makes the identification of the three categories of games impossible.

[20]Throughout the analysis, we define an option as "selfish" when it maximizes the payoff of the decision-maker. We acknowledge that in some games, choosing the "selfish" option is consistent with other types of preferences. For example, a subject who chooses the selfish option in games 1, 5 and 9 may be maximizing social welfare. We confirm that our results are robust to the exclusion of these games in regressions similar to those presented in Table 1.4. These regressions are available upon request.

Table 1.4: Determinants of Dictators' Correct Recalls

Dependent variable	*Dictator's Correct Recall*				
	All	IRA	IRAC	NIRA	IDA
	(1)	(2)	(3)	(4)	(5)
IRA treatment	Ref.	-			
IRAC treatment	-0.099***		-		
	(0.023)				
NIRA treatment	-0.015			-	
	(0.026)				
IDA treatment	0.073***				-
	(0.025)				
Selfish option	-0.057***	-0.083**	0.025	-0.014	-0.146***
	(0.017)	(0.034)	(0.025)	(0.034)	(0.035)
Dict. in disadv. position	Ref.	Ref.	Ref.	Ref.	Ref.
	-	-	-	-	-
Dict. in mixed position	-0.007	0.017	-0.039	0.009	-0.016
	(0.017)	(0.034)	(0.035)	(0.035)	(0.036)
Dict. in adv. position	-0.028	0.014	-0.025	-0.035	-0.077*
	(0.018)	(0.034)	(0.029)	(0.044)	(0.039)
Performance verbal memory	0.004**	0.0001	0.005**	0.002	0.006**
	(0.001)	(0.003)	(0.002)	(0.003)	(0.003)
Time to recall	-0.003***	-0.003	-0.001	-0.003	-0.005**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
Game order, part 1	0.002	0.006	0.002	-0.003	0.002
	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Game order, part 3	-0.005***	-0.012***	-0.004	-0.002	-0.005
	(0.002)	(0.004)	(0.004)	(0.005)	(0.004)
Age	-0.00004	-0.005**	0.010	0.001	0.001
	(0.001)	(0.003)	(0.007)	(0.001)	(0.005)
Male	0.025	0.052	0.029	-0.048	0.061*
	(0.017)	(0.033)	(0.034)	(0.036)	(0.034)
Educational attainment	0.007	-0.009	-0.006	0.010	0.021**
	(0.005)	(0.011)	(0.011)	(0.010)	(0.010)
N	3720	948	924	876	972
Clusters	310	79	77	73	81
Pseudo R^2	0.025	0.026	0.016	0.011	0.035
Log pseudolikelihood	-2050.25	-526.65	-423.81	-481.70	-593.62
Wald chi2	93.01	33.73	15.67	9.80	41.26
prob > Chi2	<0.0001	0.0002	0.1096	0.4584	<0.0001

Notes: The Table reports marginal effects from Logit regressions. Robust standard errors clustered at the individual level are in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

These models provide additional insights for our understanding of memory mechanisms. First, the participants' performance in the verbal memory task is

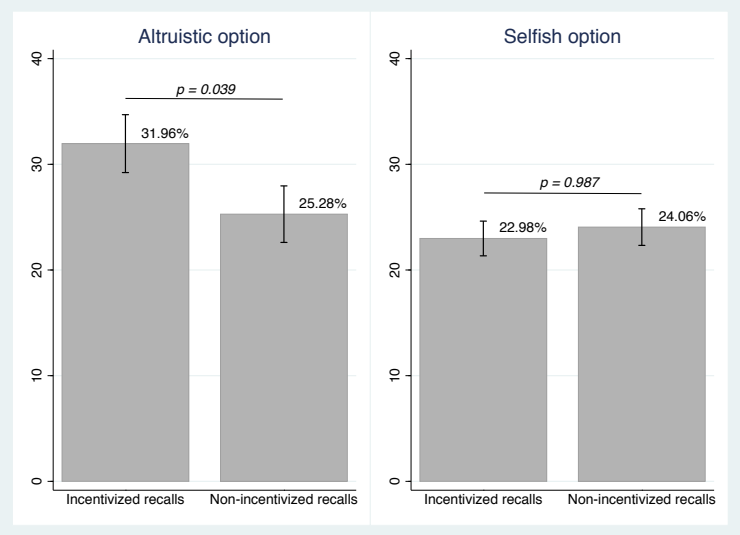
positively correlated with the likelihood of making a correct recall in part 3 (model (1)). This confirms the significant correlation between the percentage of correct recalls of the dictators in the dictator games and their performance in the verbal memory task (pairwise Pearson's correlation coefficient = 0.12, p=0.039, N=310). This effect is, however, mainly driven by the IRAC treatment (model (3)), whereas it is not observed in IRA and NIRA (models (2) and (4), respectively). In IRAC, the Pearson coefficient between the average number of correct recalls and the performance at the verbal memory task is equal to 0.21 (p=0.073). In the other treatments, it is not significant (p=0.515, p=0.246 and p=0.220 for IRA, NIRA and IDA). This gives a valuable indication that individuals actually did a memory effort to recall the amounts, but less so when they had to remember the consequences of their choices on the receiver. Second, Table 1.4 shows that spending more time to recall a given amount is negatively correlated with the likelihood of a correct recall. The extra time spent to recall does not increase accuracy. Finally, the probability of a correct recall is negatively correlated with the order in which participants had to recall this amount (p<0.001). This may be due to tiredness or weariness.

Are dictators conscious of their selective recalls? At the end of the experiment participants had to report their belief about the accuracy of their recalls on a 10-point scale, with 0 if they believe that they had no correct recall and 10 if they believe that all their recalls were correct. Pooling all the treatments, the correlation between the percentage of correct recalls and the belief about memory accuracy is highly significant (correlation coefficient=0.17, p=0.002), indicating a good perception of performance in the recall task. This correlation is highly significant for the less selfish dictators (correlation coefficient=0.21, p=0.005) but not for the more selfish ones (0.11, p=0.205). Moreover, there are differences across

treatments. The correlation is stronger when dictators are responsible for the receiver's amount (IRA, p=0.021) than when they are passive (IRAC, p=0.092). An interpretation is that in IRAC participants provided a lower memory effort and were thus more uncertain of their performance.[21]

We now introduce our second result:

Result 2 (Incentives and Selective Recalls) *Incentivizing recalls increases the percentage of dictators' correct recalls only when they chose the altruistic option.*



Notes: The Figure displays the mean percentages of dictators' correct recalls depending on the option chosen by the dictators (altruistic in the left panel, selfish in the right panel) in the IRA (incentivized recalls) and the NIRA treatments (non-incentivized recalls). p-values are from M-W tests.

Figure 1.3: Average Percentage of Correct Recalls in IRA and NIRA, by Option

[21] This interpretation is supported by an additional questionnaire on the intensity of memory effort reported on a 10-point scale. On average, the self-reported memory effort is 7.02 in IRA and 5.84 in IRAC (M-W, p<0.001). Table A2.3 in Appendix 2 displays the average reported beliefs on memory accuracy and the average reported memory effort across treatments.

Result 2 supports only partially Conjecture 2.

Support for Result 2: Each dictator was asked to recall 12 amounts, which gives 3720 (310*12) recalls in total for all treatments. 25.27% of these recalls are correct. The percentage of dictators' correct recalls is 24.43% in NIRA and 25.74% in IRA. These percentages are not significantly different (M-W, p=0.583). The picture changes when we consider the selected options separately. Figure 1.3 displays the average percentages of dictators' correct recalls in IRA and NIRA, depending on the selected option. When dictators chose the altruistic option the percentages of correct recalls is significantly higher in IRA (31.96%) than in NIRA (25.28%, M-W test, p=0.039). This is not the case when they chose the selfish option (22.98% in NIRA and 24.06% in IRA; M-W test, p=0.987). This finding shows that since dictators somewhat react to incentives, they have not completely forgotten their decisions. With incentives, dictators do not remember more their choices when they were selfish but they remember them more when these choices were altruistic. We take this as evidence that with incentives dictators provide a higher effort to recall, but they allocate this effort to retrieve selectively the memory of desirable rather than undesirable decisions.

We now turn to biased memory errors and introduce our third result:

Result 3 (Direction of Memory Errors) *Dictators are significantly more likely to over-estimate their recalls when they chose the selfish option than when chose the altruistic option. This is also the case when they bear no responsibility in the choice of option.*

Result 3 does not support Conjecture 3.

Support for Result 3: If dictators bias their memory self-servingly for self-image reasons, they should more frequently over- than under-estimate the amount given to the receiver when they make memory errors. In IRA, when dictators make an error they over-estimate the receiver's amount 48.30% of the time, regardless of their actual choices. This percentage is not significantly different from 50% (one-sample test of proportion, p=0.366). Conditioning the percentage of over-estimated recalls on decisions reveals interesting differences. On average, when they make an error dictators over-estimate the receiver's amount 31.82% of the time when they chose the altruistic option and 54.74% of the time when they chose the selfish option. The difference is significant (W test, p<0.001). However, these percentages are similar in IRAC: dictators over-estimate the receiver's amount 23.96% of the time when the program selected the altruistic option and 62.44% of the time when it selected the selfish option. The difference is also significant (W test, p<0.001) and it can hardly be motivated by the willingness to bias recalls for self-image reasons since it is common knowledge that dictators are passive.

Table A2.5 in Appendix 2 reports the marginal effects from Logit regressions in which the dependent variable is the likelihood of observing an over-estimated recall rather than an under-estimated recall, conditional on making an incorrect recall.[22] Model (1) pools all the treatments together while models (2) to (5) consider each treatment separately. The independent variables are the same as in Table 1.4. Robust standards errors are clustered at the individual level. Model

[22]We also considered two-step Heckman models, estimating first the likelihood of making an incorrect recall and then, the likelihood of over-estimating the amount given to the receiver, conditional on making a memory error. We used probit models to estimate both the selection and the outcome equations. Since the Inverse of the Mill's Ratio was significant in no model, showing that we do not need to correct for a possible selection bias, and since the results on the main variables were not affected, we omit reporting these regressions.

(1) confirms that the probability to over-estimate rather than under-estimate the receiver's amount is significantly higher when the selfish option has been selected ($p<0.001$). However, this is independent from the responsibility of the action itself since this is found not only in IRA (model (2)) but also in IRAC (model (3)). Had the dictators motivated their memory to appear more pro-social to themselves, the difference in the percentage of over-estimated recalls between the selfish and the altruistic options should have been higher in IRA than in IRAC. These findings suggest that the difference between the percentages of over-estimated amounts in the altruistic and selfish options in both treatments results more from the structure of the games (*i.e.*, lower amounts are structurally more likely to be over-estimated) than from behavioral determinants.[23]

We introduce our last result about biased memory errors based on the analysis of the magnitude of these errors, defined by their absolute value:

Result 4 (Magnitude of Memory Errors) *The magnitude of over-estimated recalls is not significantly different between altruistic and selfish choices. This is observed regardless of the dictator's responsibility for the receiver's amount.*

Result 4 rejects Conjecture 4.

Support for Result 4: Table 1.3 displays the average absolute value of memory errors across options, conditional on making an error.[24] In IRA, the average magnitude of dictators' memory errors is 5.06 when they chose the altruistic option

[23]Incidentally, the fact that in IRAC dictators are also more likely to over-estimate the receiver's amount when the program has selected the selfish option indicates that selective recalls are not driven by a concern for social-image independent from memory biases. This over-estimation cannot be explained by the willingness to appear more generous in the experimenter's eyes since in IRAC it is common knowledge that the receivers' amounts are randomly selected by the program without any intervention of the dictators.

[24]Considering the average memory error in non-absolute instead of absolute values does not qualitatively change the results.

and 5.75 when they chose the selfish one. The difference is not significant (W test, p=0.665). In IRAC, the average magnitude of memory errors is 7.27 and 7.02, respectively, and the difference is not significant either (W test, p=0.566). Further support is provided by Table A2.6 in Appendix 2 that reports the marginal effects from Tobit regressions in which the dependent variable is the absolute value of the magnitude of memory errors, conditional on making an error. The independent variables are the same as in the previous regression Tables. Tobit models are justified since data are censored on the left. Robust standard errors are clustered at the individual level. With the exception of model (4) for NIRA, models (1) to (5) show that having to recall the outcome of the selfish option has no significant impact on the magnitude of memory errors compared to when the altruistic option has been selected. Thus, when they do not recall the amount given to the receiver, dictators do not inflate their recalls self-servingly. Model (1) also indicates that the magnitude of memory errors is significantly higher when the set of available options puts the dictator in a disadvantageous position, and that a higher performance at the verbal memory task decreases the magnitude of memory errors.

1.6 Robustness Tests

This section presents three checks. We first examine whether memory errors differ from pure noise. Then, we test whether selective recalls are driven by a higher attention paid to the receiver's amount by other-regarding dictators. We finally investigate the role of guilt.

1.6.1 Memory Errors or Noise?

The recollection task was hard for the players because of the number of values to recall (12). Could the higher (lower, respectively) probability to over-estimate

the receivers' outcome when the selfish (altruistic, resp.) option has been chosen derive from the fact that dictators simply recall the average outcome of the two options? To investigate whether recalls differ from pure noise, we simulated three distributions of recalls and tested whether our results differ from these simulated distributions. The first two simulated sets of recalls follow a normal distribution centered at 18 (the mean actual receiver's amount) with a standard deviation of 4 or 2 (to simulate players that almost always reported the average receiver's amount). The third simulated set follows a uniform distribution over the range of possible recalls from 0 to 38.

This exercise reported in Table A2.7 in Appendix 2 shows that the percentage of correct recalls is significantly higher in the experimental data than in any simulated distribution (W tests, $p<0.001$). Thus, participants used their memory actively. A test of normality shows that participants did not simply report the average receiver's amount (skewness/kurtosis test for normality, $p<0.001$). Moreover, the magnitude of memory errors is significantly lower in the experimental data than in any simulated distribution (W tests, $p<0.001$). In contrast, the probability to over-estimate the receiver's amount is not significantly different between actual and simulated data (except in the second simulation). Thus, Result 3 would have been obtained for normal or uniform distributions of recalls, confirming that the difference between the probability to over-estimate a selfish *vs.* an altruistic option does not result from motivated memory, but probably from the structure of the amounts themselves.

1.6.2 Memory and Attention

In the treatments in which players had to recall the receiver's amounts, the higher percentage of dictators' correct recalls when they chose the altruistic option could

be explained not only by motivated memory but also by a higher attention paid to the receiver's amount. In contrast, when they made their decisions selfish dictators may have simply compared their own amount in the two options and ignored the receiver's amounts, leading to more memory errors. Analyzing behavior in the IDA treatment where players have to recall the amount kept by the dictator is informative because both other-regarding *and* selfish dictators are likely to have paid attention to their own amount. If the difference in recalls observed in the main treatment is driven by differential attention, we should observe no difference in recalls in IDA. In fact, in IDA also the percentage of correct recalls differs significantly between the altruistic and the selfish options (42.36% and 28.27%, respectively; W test, p=0.010, see Table 1.3). It would not be the case if recalls were only driven by differing attention according to the chosen option. Moreover, model (5) in Table 1.4 shows that having to recall the choice of the selfish option decreases significantly (at the 1% level) the likelihood of a correct recall of one's amount by the dictators. These findings support the interpretation of behavior in terms of motivated memory rather than in terms of differences in attention in our main treatments.

1.6.3 Memory and Guilt

Impression management may depend not only on the chosen option but also on the very nature of the individual. A selfish dictator who accepts his egotist nature may feel no need to recall selectively. Motivated memory may be needed only by individuals who suffer from a dissonance between their actions and their self-image, in particular those who suffer from guilt. In the post-experimental questionnaire dictators were asked to report on a 10-point scale their feelings toward the receivers, from 0 for very guilty to 10 for perfectly serene (mean=7.21, S.D=2.46, see Table A2.3 in Appendix 2). Table A2.8 in Appendix 2 displays the average

percentage of dictators' correct recalls depending on their reported feeling toward the receiver. It shows that dictators who report a feeling below or equal to the median (7) on the serenity scale exhibit a lower percentage of correct recalls than dictators who report a serenity level above the median (M-W test, p=0.028, all treatments pooled). Considering only treatments in which dictators bear responsibility in the choice of options (IRA, NIRA and IDA), more guilty dictators have also a significantly lower percentage of correct recalls than more serene dictators (M-W test, p=0.005). This is not the case when dictators bear no responsibility in the choice of options (IRAC, p=0.717). Overall, this suggests that dictators who experienced more discomfort *vis-à-vis* the receivers retrieve more selectively their recalls when they are responsible for the receiver's payoff.

1.7 Conclusion

Individuals develop a variety of deceptive strategies to maintain their self-concept when behaving in ways that may threaten their self-image, including strategic ignorance of information or delegation of decisions. In this study, we explored whether individuals manipulate their memory to appear more pro-social to themselves than they actually are. In our experiment, participants played binary dictator games and then, had to recall the amounts allocated to the receivers. This design allowed us to investigate whether dictators exhibit selective recalls and bias their memory errors self-servingly (over-estimating more often and to a larger extent the receivers' amounts), after making selfish rather than altruistic decisions.

We found evidence of *selective memory*. Individuals remember better the amount allocated to the receiver when they made altruistic rather than selfish

decisions. We interpret these asymmetric recalls as a self-deception strategy motivated by self-image concerns. This finding is consistent with previous theoretical and empirical studies on motivated memory revealing an asymmetric recall of feedback depending on whether individuals receive good or bad news about their relative performance (Bénabou and Tirole, 2002; Gottlieb, 2014; Li, 2017; Chew et al., 2018; Zimmermann, 2019). More generally, it contributes to the literature showing that individuals have motivated cognitive limitations even in the absence of risk and uncertainty (Exley and Kessler, 2018), selective memory being one of these self-serving biases. We complement the previous studies on motivated memory by showing that individuals also use selective memory in social interactions and by revealing the crucial role of personal responsibility in this process. Indeed, the asymmetric recalls that we identified are no longer observed when decisions are made at random by a robot. Moreover, our study shows that incentivizing correct recalls increases the percentage of dictators' correct recalls when they chose the altruistic option but has no effect when they chose the selfish option. This suggests that when dictators are given a monetary incentive to provide a memory effort, they allocate this effort to retrieve the memory of desirable rather than undesirable information in terms of image. Like Zimmermann (2019), we interpret the fact that incentives generate more accurate recalls as evidence against complete forgetting. Individuals selectively suppress bad news (in the case of Zimmermann, 2019) or selectively retrieve good news (in our case).

In contrast, we found no clear evidence of *biased memory errors*. Dictators are more likely to over-estimate than under-estimate the amount transferred to the receiver after choosing the selfish rather than the altruistic option. But this does not prove the existence of motivated memory since this also applies when dictators are not responsible for the amount transferred to the receiver. Moreover, the

magnitude of memory errors is not significantly different across options. Thus, individuals recall selectively but they do not manipulate their memory self-servingly to appear altruistic when they were selfish. There are several possible explanations for the absence of biased memory errors. First, dictators may not bias their memory because it is common knowledge that the experimenter knows the information dictators are asked to recall. In a different domain, it has been shown that the propensity of individuals to lie differs depending on whether the experimenter can or cannot observe the truth (Gneezy et al., 2018). The same might apply to our setting: forgetting is unverifiable but inflating one's recalls systematically is detectable. An extension of our study could be to design games in which participants know that the experimenter cannot observe memory errors at the individual level. Second, the limited bias of memory errors may also result from the short span of time between action and its recollection. We chose to hold the action and recollection phases in the same session to make it cognitively doable for the subjects to retrieve their memory. But it is possible that a larger span of time is needed to bias recalls self-servingly. A natural extension would be to vary the length between the decision and the recollection phases to test how it affects biased memory errors.

Other possible extensions can be thought of to study biases in memory errors. Even if a majority of individuals probably prefer to think of themselves as generous rather than egoist and unfair, the dissonance between making selfish decisions in dictator games when a pro-social alternative is available and maintaining a positive self-image may not be strong enough to generate an internal conflict. Introducing decisions that threaten self-image more deeply, by revealing to participants a more precise and valuable information about their intrinsic nature, could generate a stronger need for biased memory. Finally, in our design individuals could manipulate their memory only for hedonic reasons. Another interesting extension

would be to introduce strategic reasons to use selective memory and to bias recalls asymmetrically. Testing how memory can be manipulated self-servingly for motivational purposes is left for further investigation.

Appendices

A1: Instructions (*translated from French*)

Introduction

We thank you for participating in this experiment on decision-making. Please switch off your cellphone and put it away. You are not allowed to communicate with the other participants. If you have any question during the session, you can press the red button on the side of your cubicle. An experimenter will come and answer to your questions in private. During the session, you will have to make several decisions. These decisions are anonymous and can earn you money. Regardless of these decisions, you will receive a five euros show-up fee. Your earnings will be expressed in Experimental Currency Units (ECU) and converted into Euros at the following rate: 4 ECU = €1. You will be paid in cash and in private, in a separate room. Other participants will not be informed of your earnings.

The session consists of 4 parts. At the end of each part, you will receive the instructions for the next part. All the instructions will be displayed on the screen.

Please read again these instructions. If you have any questions, please raise your hand or press the red button. When you are ready, press OK to see the instructions for Part 1.

Instructions Part 1

This part consists in 12 independent periods. At the beginning of the part, you will be assigned a role, either A or B. You will keep this role for the 12 periods.

At the beginning of each period, you are going to be randomly matched with another participant, to form a pair. In each pair, a participant has the role A and the other has the role B. If you have the role A, you are matched with a participant with role B and if you have the role B, you are matched with a participant with role A. Participant B has no decision to take.

The decision of participant A consists in choosing the preferred option between two options: option X and option Y. Each option is composed of two amounts: the first amount corresponds to the payoff of participant A, the second amount corresponds to the payoff of participant B.

To validate his choice, participant A has to click on the option he prefers and type the amounts corresponding to that option in the corresponding box. It is very important to look carefully at the two amounts of each option before choosing the preferred option. Once A has chosen his preferred option, B is informed of the option chosen by A. Player B has in turn to click on the option chosen by A and type the amounts corresponding to this option in a box. Then, a new pair is formed and a new period starts.

How is determined your payoff in this part?
At the end of the session, the program selects at random one period among the twelve. Participant A receives the first amount corresponding to the option he has chosen in this period. Participant B receives the second amount corresponding to the option chosen by participant A in this period. For example, if the option chosen by A in the randomly selected period is (20, 12): A receives 20 ECU and B receives 12 ECU.

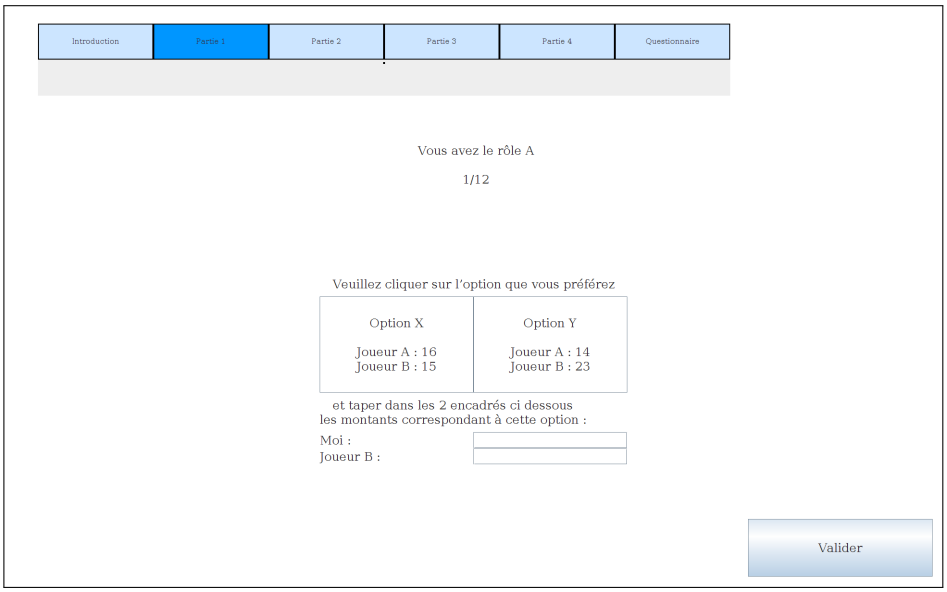
Please read again these instructions. If you have any questions, raise your hand or press the red button. Before starting this part, you have to answer to an understanding questionnaire. Press OK to answer to these questions.

Instructions for Part 2 (displayed on the subjects' screen after completing Part 1)

In this part, you have 8 minutes to solve mazes. There are 30 mazes in total with different levels of difficulty (10 easy, 10 intermediate, 10 difficult). You can skip a maze, but you cannot return to a previous maze. To solve a maze, you have to move a small character from the top left of the maze to the exit, at the bottom right of the maze. To move the character, use the left, right, top and down arrows of your keyboard. Before starting this 8-minute part, you will have the opportunity to practice on a maze. Solving this practice maze is not paid.

How is determined your payoff in this part?
You will earn 1 ECU for each maze solved.

Please read again these instructions. If you have any questions, raise your hand or press the red button. When you are ready, press OK to start Part 3.



Translation: "You have role A. Please click on the option you prefer and type in the amounts corresponding to this option in the two boxes below. Me. Player B."

Figure A1.1: Example of a screen in Part 1, player A

Figure A1.2: Example of a maze in Part 2

Instructions for Part 3 (displayed on the subjects' screen after completing Part 2)

In each of the twelve games in Part 1, you (respectively, player A) had to choose the option you (respectively, he) preferred among two. Each option contained two

amounts: the first amount corresponded to your (respectively, player A's) payoff and the second amount corresponded to the payoff of player B (respectively, your payoff). The amounts between you (respectively, player A) and player B (respectively, A) were different between the two options.

You are going to see again, successively and in a random order, the options that you have seen in each of the 12 periods of Part 1. However, in the option you (respectively, player A) have (has) chosen, the amount received by player B (respectively, you) will be hidden and replaced by a question mark, as in the example below. Your task consists in recalling this amount. In the above example, if you (respectively, player A) have (has) chosen option X that gave you (respectively, player A) 20 ECU, you have to recall the amount replaced by the question mark. This amount corresponds to player B's (respectively, your) payoff in the option you (respectively, player A) have chosen. Note that the amounts are bounded between 0 and 38. This means that no amount can be lower than 0 and higher than 38.

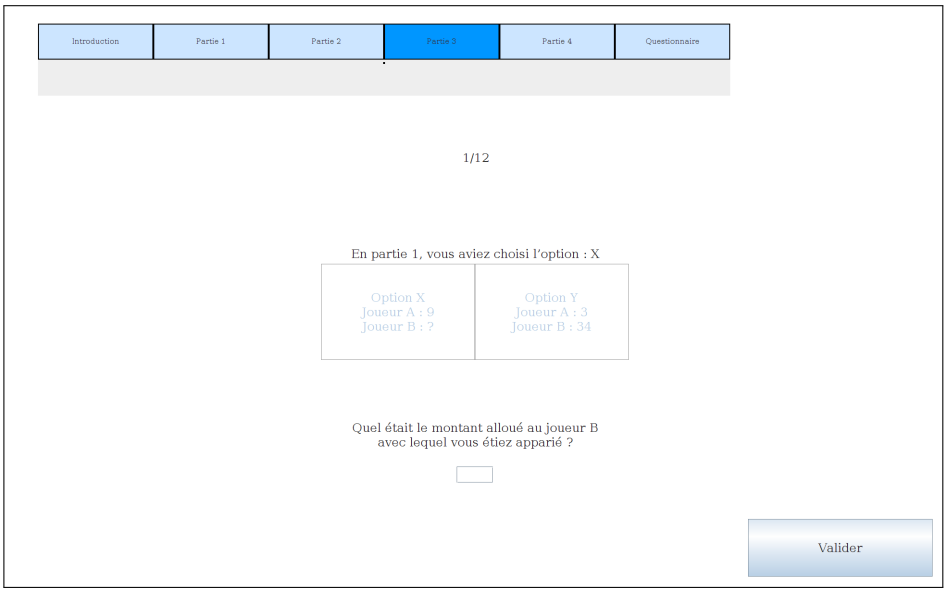
How is determined your payoff in this part?
At the end of the session, two recalls will be randomly selected. Your payoff depends on the accuracy of your recall in each of these two recalls. If your recall is correct, you will earn 8 ECU (€2). If your recall is correct plus or minus one unit, you will earn 4 ECU (€1). For example, if the amount to recall is 24 and that your recall is 24, you earn 8 ECU. If your recall is 23 or 25, you earn 4 ECU. If your recall is lower that 23 or higher than 25, you do not earn anything. You will be informed of your total number of correct recalls at the end of the session.

Please read again these instructions. If you have any questions, raise your hand or press the red button. When you are ready, please press OK to start Part 3.

Instructions for Part 4 (displayed on the subjects' screen after completing Part 3)

This part consists in 3 independent rounds. In each round, you will see a list of 15 words corresponding to singular nouns, without accent and written in lowercase. Each word will be displayed on your screen one by one during a few seconds. Your task consists in memorizing these words. Once you will have watched the 15 words, you will have to type the highest number of words that you recall from the list in a dedicated box. You will have 2 minutes to write the words you recall. The order in which you recall the words does not matter.

How is determined your payoff in this part?
At the end of the session, one round out of the three will be randomly selected. For each word correctly recalled in that round, you will earn 1 ECU.

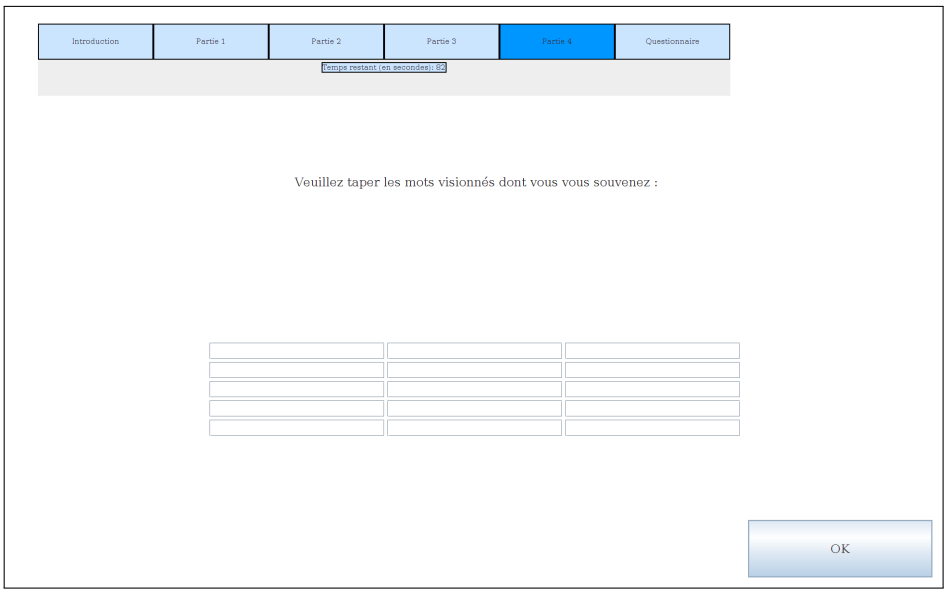


Translation: "In part 1, you have chosen option X. What was the amount allocated to the Player B you were matched with?"

Figure A1.3: Example of a screen in Part 3, player A

Please read again these instructions. If you have any questions, raise your hand or press the red button. When you are ready, press OK to start Part 4.

A2: Tables



Translation: "Please type in the words that you remember."

Figure A1.4: Example of a screen in Part 4

Table A2.1: Summary Statistics - Decisions in the Dictator Games

Games	Option X	Option Y	Percent. of dictators choosing opt. Y				
	Altruistic	Selfish	All	IRA	IRAC	NIRA	IDA
1	(2, 32)	(10, 30)	86.77	97.47	50.65	98.63	100.00
2	(3, 34)	(9, 28)	87.42	98.73	54.55	95.89	100.00
3	(5, 35)	(7, 27)	75.81	75.96	62.34	82.19	82.72
4	(6, 36)	(6, 26)	28.06	21.52	50.65	26.03	14.81
5	(11, 20)	(19, 18)	88.06	98.73	54.55	98.63	100.00
6	(12, 22)	(18, 16)	89.35	98.73	61.03	98.63	98.77
7	(14, 23)	(16, 15)	73.87	83.54	48.05	84.93	79.01
8	(15, 24)	(15, 14)	31.61	24.05	58.44	27.30	17.28
9	(20, 8)	(28, 6)	81.94	91.14	51.95	91.78	92.59
10	(21, 10)	(27, 4)	69.03	77.22	53.25	72.60	72.84
11	(23, 11)	(25, 3)	49.35	54.43	45.45	52.05	45.68
12	(24, 12)	(24, 2)	19.68	10.13	51.95	8.22	8.64
Total			65.08	69.30	53.57	69.75	67.70

Notes: The first numbers in parentheses in columns 2 and 3 indicate the dictator's amounts, and the second numbers indicate the receiver's amounts. The percentages of dictators choosing option Y are significantly different neither between IRA and IDA, nor between IRA and NIRA. The percentages of option Y selected randomly by the program (IRAC treatment) are always significantly different from the percentages of dictators choosing option Y (treatment IRA) at 5% level, except for games 3 (Mann-Whitney tests, p=0.066) and 11 (p=0.264).

Table A2.2: Summary Statistics - Participants, by treatment

Treatments	All	IRA	IRAC	NIRA	IDA
Male	47.23%	43.67%	50.00%	45.20%	51.23%
Age	22.55	22.84	21.06***	24.64	21.62**
Number of participants	578	158	154	146	162
Number of sessions	24	7	6	6	7
Ave. num. of part. per session	24.51	22.57	25.67	24.33	23.14

Notes: The Table reports the results of two-tailed M-W tests in which each individual is taken as an individual observation. NIRA, IDA and IRAC are compared to IRA.

Table A2.3: Summary Statistics on Each Part, by Treatment

		All	IRA	IRAC	NIRA	IDA
Part 1	Percentage of selfish choices (out of 12)	65.08	69.30	53.57***	69.75	67.70
Part 2	Num. of solved mazes	12.21	12.08	12.45**	11.64	12.64*
Part 3	Num. of correct recalls (out of 12)	3.01	2.93	2.30***	2.87	3.88***
Part 4	Num. of correct words (out of 45)	24.73	25.31	24.98	23.78**	24.78
Quest.	Reported belief on mem. accu. (0-10 scale)	4.16	4.44	3.54***	4.21	4.43
	Reported memory effort (0-10 scale)	6.50	7.02	5.84***	6.51	6.61*
	Reported feeling toward the other player:					
	Dictator (0: very guilty; 10: very serene)	7.21	7.14	7.45	6.89	7.32
	Receiver (0: very angry; 10: very serene)	6.45	6.61	5.91**	6.41	6.85

Notes: The Table reports the results of two-tailed M-W tests in which each individual is taken as an individual observation. IRAC, NIRA, and IDA treatments are compared to IRA. mem. accu. = memory accuracy.

Table A2.4: Percentage of Dictators' Correct Recalls, by Option and by Position (IRA)

	Position			*p-values*	
	Disadv. (1)	Mixed (2)	Adv. (3)	(1)-(2)	(1)-(3)
Altruistic	35.71%	28.00%	31.82%	*0.165*	*0.873*
Selfish	21.12%	25.31%	22.28%	*0.226*	*0.872*

Notes: The p-values are from two-tailed W tests in which each individual gives one independent observation.

Table A2.5: Determinants of Dictators' Over-Estimated Recalls

Dependent variable	*Dictator's Overestimated Recall*				
	All	IRA	IRAC	NIRA	IDA
	(1)	(2)	(3)	(4)	(5)
IRA treatment	ref.	-			
IRAC treatment	0.024		-		
	(0.025)				
NIRA treatment	-0.007			-	
	(0.028)				
IDA treatment	-0.054				-
	(0.034)				
Selfish option	0.247***	0.312***	0.362***	0.360***	-0.164***
	(0.023)	(0.048)	(0.026)	(0.039)	(0.045)
Dict. in disadv. position	ref.	ref.	ref.	ref.	ref.
	-	-	-	-	-
Dict. in mixed position	0.139***	0.271***	0.228***	0.150***	-0.115***
	(0.022)	(0.035)	(0.035)	(0.040)	(0.039)
Dict. in disadv. position	0.352***	0.532***	0.460***	0.477***	-0.183***
	(0.028)	(0.027)	(0.034)	(0.045)	(0.046)
Performance verbal memory	0.002	-0.0002	0.006**	0.003	-0.002
	(0.002)	(0.004)	(0.002)	(0.003)	(0.005)
Time to recall	0.002**	0.001	-0.002	-0.0003	0.006***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Game order, Part 1	0.0004	-0.001	0.001	0.001	-0.001
	(0.002)	(0.004)	(0.004)	(0.005)	(0.005)
Game order, Part 3	0.010***	0.004	0.0001	0.0002	0.014***
	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
Age	0.004**	0.003	-0.001	0.002	0.014***
	(0.002)	(0.003)	(0.010)	(0.002)	(0.004)
Male	-0.036*	-0.007	-0.047*	-0.052	-0.039
	(0.021)	(0.043)	(0.028)	(0.041)	(0.053)
Educational attainment	-0.010*	-0.011	-0.012	-0.009	-0.015
	(0.006)	(0.016)	(0.014)	(0.010)	(0.010)
N	2780	704	761	662	653
Clusters	310	79	77	73	81
Pseudo R^2	0.1083	0.2181	0.2753	0.1780	0.0672
Log pseudolikelihood	-1709.08	-381.22	-378.65	-376.76	-415.74
Wald chi2	244.32	167.24	235.53	121.73	48.42
Prob > chi2	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Notes: Marginal effects from Logit models are reported, with robust standard errors clustered at the individual level in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A2.6: Determinants of Dictators' Magnitude of Memory Errors

Dependent variable	*Dictator's Magnitude of Memory Error*				
	All	IRA	IRAC	NIRA	IDA
	(1)	(2)	(3)	(4)	(5)
IRA treatment	ref.	-			
IRAC treatment	1.414***		-		
	(0.380)				
NIRA treatment	0.010			-	
	(0.369)				
IDA treatment	-1.637***				-
	(0.338)				
Selfish	-0.278	-0.126	-0.078	-0.927*	0.258
	(0.245)	(0.514)	(0.429)	(0.541)	(0.389)
Dict. in disadv. position	ref.	ref.	ref.	ref.	ref.
	-	-	-	-	-
Dict. in mixed position	-1.186***	-0.862	-1.368**	-1.740***	-0.705
	(0.286)	(0.547)	(0.598)	(0.639)	(0.483)
Dict. in adv. position	-1.269***	-1.449**	-1.861***	-1.649**	0.231
	(0.318)	(0.574)	(0.618)	(0.790)	(0.548)
Performance verbal memory	-0.062**	-0.088**	-0.019	-0.107*	-0.041
	(0.026)	(0.044)	(0.051)	(0.064)	(0.042)
Time to recall	0.0001	-0.044**	0.014	0.009	0.029
	(0.012)	(0.019)	(0.023)	(0.035)	(0.019)
Game order, Part 1	-0.002	-0.066	0.016	-0.008	0.028
	(0.026)	(0.058)	(0.052)	(0.056)	(0.035)
Game order, Part 3	-0.096***	-0.068	-0.051	-0.188**	-0.032
	(0.030)	(0.061)	(0.065)	(0.076)	(0.038)
Age	0.063***	0.024	0.228	0.045	0.152**
	(0.022)	(0.045)	(0.153)	(0.027)	(0.071)
Male	-0.902***	-1.337***	-0.959*	-0.401	-0.941**
	(0.257)	(0.489)	(0.541)	(0.518)	(0.467)
Educational attainment	0.052	0.106	0.052	0.180	-0.206*
	(0.070)	(0.145)	(0.152)	(0.158)	(0.116)
N	2780	704	761	662	653
Clusters	310	79	77	73	81
Pseudo R^2	0.0128	0.0079	0.0053	0.0116	0.0170
Log pseudolikelihood	-8599.61	-2161.95	-2437.23	-2100.28	-1804.99
F	13.30	3.68	2.21	3.82	3.08
$p > F$	<0.0001	0.0001	0.0157	<0.0001	0.0008

Notes: Marginal effects from Tobit models are reported with robust standard errors clustered at the individual level in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A2.7: Dictators' Recalls, Actual and Simulated Distributions

	Data	Normal (s.d.=4)	Normal (s.d.=2)	Uniform
Percentage of correct recalls	22.49	3.48***	2.09***	3.88***
Percentage of over-estimated recalls	46.81	47.15	50.15***	47.92
Magnitude of memory errors	6.09	9.16***	8.76***	12.31***
Clusters	458	458	458	458

Notes: The first simulated set of recalls follows a normal distribution centered at 18 (the mean actual receiver's amount) with a standard deviation of 4. The second simulated set of recalls follows a normal distribution centered at 18 but with a standard deviation of 2 to simulate players that may have almost always reported the average receiver's amount. The third simulated set of recalls follows a uniform distribution over the range of possible recalls from 0 to 38. For each distribution, three variables have been computed: a binary variable equal to 1 if the recall is correct and 0 otherwise, a binary variable equal to 1 if the recall is overestimated and 0 otherwise, and a variable that indicates the magnitude of errors and is equal to the difference between the simulated recall and the actual amount. *p*-values from W tests indicate whether each simulated distribution differs from the actual results. Each individual gives one independent observation. *** $p<0.01$.

Table A2.8: Average Percentage of Dictators' Correct Recalls Depending on Their Reported Feeling Toward the Receiver

	Reported Feeling		
	More Serene	More Guilty	*p-value*
All	26.44% (214)	22.66% (96)	0.028
Dictators **responsible** for the decision	29.63% (162)	23.59% (71)	0.005
Dictators **not responsible** for the decision	16.51% (52)	20.00% (25)	0.717

Notes: Dictators had to report on a 10-level scale their feeling toward the receiver, from 0 (very guilty) to 10 (very serene), inclusive. The reported guilty group includes dictators reporting a value lower or equal to 7 (the median of reported feeling); the reported serene group includes dictators reporting a value higher than 7. *p*-values from M-W tests are in italics. The average number of correct recalls of each individual gives one independent observation.

A3: Figure

Notes: Each dot represents one recall. Each dot on the diagonal represents an amount recalled accurately. For a better view, we used the "jitter" option in Stata that differentiates dots located in the same position.

Figure A3.5: Recalled and Actual Amounts in the Dictator Games, by Treatment

A4: Alternative Definitions of Correct Recalls

In Table A4.9 a recall is defined as correct if the recalled amount is exactly equal to the actual amount. In Table A4.10 a recall is defined as correct if the recalled amount is equal to the actual amount plus or minus two units.

A5: Analysis of the Receivers' Recalls

In our experiment, participants play 12 binary dictator games. Then, after performing a distraction task, they are asked to recall the amounts allocated to the receivers. While the Results section only reports the dictators' recalls (comparing dictators' and receivers' recalls cannot provide a clean identification of dictators'

Table A4.9: Determinants of Dictators' Correct Recalls (+/- 0 units)

Dependent variable	*Dictator's Correct Recall*				
	All	IRA	IRAC	NIRA	IDA
	(1)	(2)	(3)	(4)	(5)
IRA treatment	ref.	-			
IRAC treatment	-0.068***		-		
	(0.0198)				
NIRA treatment	-0.012			-	
	(0.022)				
IDA treatment	0.080***				-
	(0.022)				
Selfish option	-0.075***	-0.060*	-0.010	-0.016	-0.198***
	(0.015)	(0.032)	(0.022)	(0.028)	(0.029)
Dict. in disadv. position	ref.	ref.	ref.	ref.	ref.
Dict. in mixed position	0.030**	0.053*	-0.035	0.051**	0.043
	(0.014)	(0.029)	(0.028)	(0.024)	(0.030)
Dict. in adv. position	-0.036**	-0.003	-0.023	-0.030	-0.090**
	(0.014)	(0.031)	(0.025)	(0.033)	(0.036)
Performance verbal memory	0.004***	0.003	0.005**	0.001	0.005**
	(0.001)	(0.003)	(0.002)	(0.002)	(0.003)
Time to recall	-0.003***	-0.002	-0.0001	-0.004**	-0.005**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
Game order, part 1	0.001	0.006**	-0.002	-0.0004	-0.002
	(0.002)	(0.003)	(0.003)	(0.003)	(0.004)
Game order, part 3	-0.008***	-0.011***	-0.007**	-0.008**	-0.008**
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)
Age	0.0004	-0.002	0.004	0.001	0.0002
	(0.001)	(0.002)	(0.006)	(0.001)	(0.004)
Male	0.011	0.037	0.005	-0.056*	0.043
	(0.015)	(0.030)	(0.031)	(0.032)	(0.032)
Educational attainment	0.006	-0.004	0.003	0.007	0.016*
	(0.005)	(0.010)	(0.011)	(0.009)	(0.009)
N	3720	948	924	876	972
Clusters	310	79	77	73	81
Pseudo R^2	0.0427	0.0352	0.0279	0.0332	0.0729
Log pseudolikelihood	-1612.24	-407.41	-318.03	-363.97	-493.07
Wald chi2	120.56	32.93	20.17	29.51	79.65
Prob > chi2	<0.0001	0.0003	0.0277	0.0010	<0.0001

Notes: Marginal effects from Logit models are reported, with robust standard errors clustered at the individual level in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

motivated memory because receivers may also motivate their memory, albeit for

Table A4.10: Determinants of Dictators' Correct Recalls (+/- 2 units)

Dependent variable	*Dictator's Correct Recall*				
	All	IRA	IRAC	NIRA	IDA
	(1)	(2)	(3)	(4)	(5)
IRA treatment	ref.	-			
IRAC treatment	-0.075***		-		
	(0.025)				
NIRA treatment	0.014			-	
	(0.027)				
IDA treatment	0.105***				-
	(0.027)				
Selfish option	-0.060***	-0.098***	0.001	-0.014	-0.136***
	(0.018)	(0.036)	(0.0300)	(0.040)	(0.039)
Dict. in disadv. position	ref.	ref.	ref.	ref.	ref.
Dict. in mixed. position	-0.017	0.009	-0.030	-0.002	-0.040
	(0.019)	(0.039)	(0.042)	(0.036)	(0.035)
Dict. in adv. position	-0.005	0.048	0.052	-0.003	-0.115***
	(0.020)	(0.035)	(0.036)	(0.049)	(0.041)
Performance verbal memory	0.005***	-0.001	0.008***	0.003	0.009**
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)
Time to recall	-0.003**	-0.002	-0.001	-0.003	-0.005**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Game order, part 1	-0.0004	0.007	-0.004	-0.004	-0.001
	(0.002)	(0.005)	(0.004)	(0.005)	(0.005)
Game order, part 3	-0.003	-0.008**	-0.004	-0.003	-0.002
	(0.002)	(0.004)	(0.004)	(0.005)	(0.004)
Age	-0.0002	-0.002	0.0004	0.0003	0.0002
	(0.001)	(0.002)	(0.007)	(0.002)	(0.003)
Male	0.030	0.068**	0.056	-0.075**	0.073*
	(0.019)	(0.034)	(0.036)	(0.040)	(0.040)
Educational attainment	0.006	-0.007	-0.009	0.008	0.026**
	(0.005)	(0.010)	(0.011)	(0.011)	(0.010)
N	3720	948	924	876	972
Clusters	310	79	77	73	81
Pseudo R^2	0.0196	0.0182	0.0186	0.0101	0.0352
Log pseudolikelihood	-2393.48	-601.48	-549.56	-567.53	-645.62
Wald chi2	69.48	25.11	21.12	10.33	40.60
Prob > chi2	<0.0001	0.0051	0.0203	0.4116	<0.0001

Notes: Marginal effects from Logit models are reported, with robust standard errors clustered at the individual level in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

different reasons), this section presents a brief analysis of the receivers' recalls.

First, receivers do not exhibit selective recalls. Their percentage of correct recalls is 25.43% when the dictator has chosen the altruistic option and 22.07% when he has chosen the selfish option. The difference is not significant (p=0.383, M-W, IRA treatment). We also find no significant differences between the percentages of correct recalls when the dictator chose the altruistic *vs.* the selfish option in the IRAC and NIRA treatments (see Table A5.11). In the IDA treatment in which receivers have to recall the dictator's amount, they exhibit a higher percentage of correct recall when the dictator chose the altruistic option (39.81%) than when he chose the selfish option (28.11%, p=0.013, M-W). We find no significant difference in the rate of correct recalls between the receivers who have been more frequently exposed to selfish dictators and the other receivers, in either treatment.[25]

Second, there is no statistical difference in the percentage of correct recalls between the IRA and IRAC treatments, neither when the altruistic option was selected (p=0.174, M-W), nor when the selfish option was selected (p=0.972, M-W). We also find no evidence of statistical differences between the IRA and NIRA treatments, neither conditional on the altruistic option (p=0.181) nor conditional on the selfish option (p=0.588).

Third, receivers are significantly more likely to over-estimate their recalls when dictators chose the selfish option than when they chose the altruistic option (see Table A5.11). This could suggest that receivers bias their memory to derive positive anticipated utility from high-expected future payoffs. However, the opposite is observed in the IDA treatment in which receivers have to recall the dictator's

[25]In IRA, the receivers who have been exposed to selfish dictators more than 8 times out of 12 (N=37) exhibit the same average percentage of correct recalls (22.30%) than those who have been less exposed (N=42; 23.80%), and the difference is not significant (M-W test, p=0.522). In IRAC, the respective percentages (and numbers) are 21.63% (N=37) and 19.79% (N=40), and they are not significantly different either (p=0.409).

amount. This result suggests that the likelihood of overestimating the amount is more driven by the amount itself than by the nature of the option. Indeed, when the amount is low (selfish option in IRA, IRAC and NIRA and altruistic option in IDA), it has by construction a higher likelihood of being over-estimated. We also find no evidence of a different magnitude of memory errors between the altruistic and the selfish options (see Table A5.11).

Table A5.11: Summary Statistics - Receivers' Recalls

	IRA (1)	IRAC (2)	NIRA (3)	IDA (4)
Percentage of correct recalls, by option				
Alt. option	25.43% (291)	19.81% (429)	21.51% (265)	39.81%*** (314)
Self. option	22.07% (657)	21.41% (495)	24.22% (611)	28.11%*** (658)
p-values	*0.383*	*0.600*	*0.170*	*0.013*
Percentage of over-estimation, by option				
Alt. option	27.65% (217)	27.33% (344)	30.77% (208)	64.02%*** (189)
Self. option	53.52% (512)	61.44%*** (389)	57.24% (463)	38.90%*** (473)
p-values	*<0.001*	*<0.001*	*<0.001*	*<0.001*
Magnitude of absolute memory errors, by option				
Alt. option	5.80 (291)	7.16** (429)	6.12 (265)	3.89*** (314)
Self. option	5.38 (657)	6.41** (495)	5.75 (611)	3.73*** (658)
p-values	*0.588*	*0.097*	*0.214*	*0.932*

Notes: The *p*-values in lines are from M-W tests and those in columns (altruistic *vs.* selfish option) are from W tests. Each individual gives one independent observation. Numbers in parentheses display the number of individual observations.

Chapter 2

Mood-driven or goal-driven memory biases?[1]

2.1 Introduction

Human beings are not always good judge of themselves. Both experiments and surveys show that people tend to be unreasonably self-confident and to have an unrealistically good self-image, in particular when compared to others (Greenwald, 1980; Taylor and Brown, 1988; Gilovich, 2008; Williams and Gilovich, 2008). Over 90% of university professors think they are better at their jobs than their average colleague (Gilovich, 2008). Almost 90% of drivers think to be safer than the median driver (Svenson, 1981). 86% of American think to be more "happy and contented" than about two-thirds of the people (Lykken and Tellegen, 1996).

Over-confidence has some advantages: it can be a trigger of audacious actions, a driver of personal progress and an instrument for self-regulation. However, it can also lead to tainted economic and financial decisions. Start-up founders pre-

[1]This chapter is a joint work with Alberto Prati, PhD candidate at Aix-Marseille School of Economics.

dict their profits to be positive, even when they correctly predict that most new companies fail (Camerer and Lovallo, 1999). As a consequence, the economy is flooded with an excess of new businesses with respect to market capacity. CEOs often consider the market valuation of their companies to be lower than their actual value (Malmendier and Tate, 2005), so that they overinvest when they have abundant internal funds, but suboptimally restrict investments when they require market financing. Traders tend to evaluate their private information as better than average (Odean, 1998), thus leading to an inflated trading volume.

Why does overconfidence happen? In the debate on the origins of the phenomenon, memory stands out as a key factor.[2] At least since the pioneering works of Tversky and Kahneman (1973),[3] memory is known to be imperfect and biased in some predictable ways, so that selective retrieval of previously acquired information can explain why economic agents delude themselves into thinking to be better than average. If negative feedback tends to be forgotten and positive feedback to be recalled, the overall self-image ends up being positively inflated. Although some recent studies confirm that individuals recall more accurately positive than negative feedback (Chew et al., 2018; Zimmermann, 2019; Li, 2017), the origins of the phenomenon have not been clearly defined.

Two main explanations of this recall asymmetry have emerged. On the one hand, theories of motivated memory highlight self-serving explanations, according

[2]Some prominent explanations of over-confidence include the hard-easy effect (Fischhoff et al., 1977), confirmation bias (Koriat et al., 1980), self-attribution bias (Gervais and Odean, 2001) and self-signalling (Mijović-Prelec and Prelec, 2010). It is worth noting that all these mechanisms require either neglectful attention, biased information processing or biased memory as mediators for the initial feedback to be distorted. In our experiment, we will provide an attention check and elicit both ex-ante and ex-post beliefs so as to leave memory as the only channel of information distortion. For a thorough discussion of overconfidence in economics and finance see, Skala (2008) and the symposium on the *Journal of Economic Perspectives*, Vol. 29 No. 4 Fall 2015.

[3]Some forerunner evidence of systematic recall errors can be found in Bartlett (1932) and Von Restorff (1933).

to which people recall positive information to enhance or protect themselves (self-enhancement effect). On the other hand, theories of associative memory explain asymmetric recall as the result of the enhanced accessibility of positive information and the attenuated accessibility of negative information, when people are in a non-negative mood (mood-congruency effect). While the first theory depicts recall errors as mostly goal-driven behaviors toward well-being, the second one conceives recall errors as unintentional consequences of high affective well-being.

This study is the first attempt in economics to disentangle two forces which have been proposed as explanations of memory failures for self-relevant information. Our design allows us to create controlled situations where self-enhancement and mood-congruency effects predict different outcomes. Although the two principles are not mutually exclusive and most existing evidence is consistent with both theories, understanding the driving force of the phenomenon is crucial to predict some suboptimal economic decisions and to develop policies aimed at mitigating or removing overconfident behaviors, whether necessary.

Let us take the example of an investor who owns a stock for two periods and decides whether to buy or sell additional stocks at the end of the second period. Insofar as we acknowledge investor's memory to be biased, we can rationalize sub-optimal future investment and violations of Bayesian updating. If the market value of the asset initially decreases and subsequently increases, both self-enhancement and mood-congruency predict some over-investment. Consider instead the case where market value increases in the first period and decreases in the following period. Mood-congruency predicts negative over-reaction to the news: selective recall pushes the deceived investor to over-sell the asset (Bodoh-Creed, 2017). Self-enhancement predicts the opposite: the investor tends to forget the less preferred

outcomes, so that she over-invests in the stock (Gödker et al., 2019).

The relative dominance either of self-enhancement or of mood-congruency bears some policy implications to mitigate overconfidence. The first effect suggests the importance of correcting *ex-ante* beliefs on what is helpful for oneself. If asymmetric feedback recall is goal-driven, more cautious financial behavior is accomplished whether the investor is aware that optimistic information is not only self-enhancing, but also biased toward risk. If feedback recall is mood-driven, tainted decisions are a collateral effect of good affective states. In this case, the investor should be informed that she is more likely to take over-optimistic decisions when in positive mood.

To identify and disentangle the driving force of asymmetric feedback recall, we set a laboratory experiment where the two theories offer divergent predictions. The laboratory offers a controlled environment where recall accuracy can be carefully assessed, and mood exogenously manipulated. Based on the design by Zimmermann (2019), subjects had to perform an IQ test and received incomplete feedback about their performance relative to their peers. One month later, they came back to the laboratory and were asked to recall their feedback. Before retrieval, we intervened or not on their mood, using Andrade et al. (2015)'s procedure.

Our results provide support for the existence of self-enhancement memory. First, individuals exhibit higher percentage of correct recalls when the feedback was positive than when it was negative. Second, when they do *not* recall correctly, individuals overestimate the number of positive feedback they received. Together, these results show overly optimistic recall of past feedback and replicate the findings in Zimmermann (2019). By contrast, we do not find clear evidence

of mood-congruent memory, although our manipulation proves to be effective in inducing the desired affective state. Individuals do not exhibit a higher percentage of correct recalls when the feedback to retrieve is congruent with their mood. Overall, our results confirm the effect of self-enhancement memory as a driver of asymmetric recall, but they fail to support any role of mood-congruency.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the experimental design and procedures. Section 4 introduces the theoretical predictions of both self-enhancing memory and mood-congruent memory. Section 5 reports the results and section 6 concludes.

2.2 Related Literature

The traditional goal of memory is to accurately reproduce the maximum amount of information. This task is far from being fully accomplished. Not only everyday experience teaches that memory is limited and imperfect, but also decades of research have unveiled that memory is endogenous: recall errors are, to some extent, systematic and predictable.

When we narrow our focus on self-relevant memories, overall two dominant forces emerge: *"A tendency for individuals to recall positive information (a self-enhancement effect), particularly when in a positive mood (a mood-congruent recall effect)"* (Baumeister et al., 2001, p.344). Self-enhancing memory refers to the psychological phenomenon of remembering relevant items in a self-serving fashion, so that unfavorable information tends to be forgotten or manipulated.[4] Mood-

[4]Various authors may refer to this same process using different terms such as self-serving bias, self-protecting bias or egocentric bias. Throughout this paper, we will use the overarching adjective "self-enhancing".

congruent memory refers to the tendency to recall more accurately items which are of the same valence as the current affective state, so that people in good mood tend to better remember positive than negative information, and vice versa for people in bad mood.

Psychologists and, more recently, economists have been trying to rationalize these empirical regularities in memory errors. They have modeled imperfect memory as a strategic compromise between conflicting goals, where selective recall and confabulation can be optimal outcomes. Although these models are diverse and multiform, we can broadly distinguish two different approaches, according to which effect they focus on and which conflicting goals are at stake. Self-enhancement effects highlight the conflict between exact information and the demand for self-esteem. Mood-congruency highlights the contrast between exact information and effort reduction. In the following paragraphs, we will review some cross-disciplinary relevant contributions to the understanding of the two phenomena.

2.2.1 Self-enhancing Memory

Psychologists generally agree that memory contributes to maintain a high self-image. Taylor (1991)'s influential theory of mobilization-minimization tries to conciliate one's motivation to prioritize positive image with the antagonistic weight of negative information. Taylor describes two functionally distinct processes. At first, negative signals are generally more salient, as they are potentially threatening and request a quick response (mobilization process). However, in a second stage, information is reviewed in a self-enhancing way and attention is focused toward its desirable aspects (minimization process). In the case of self-relevant feedback, Taylor's theory predicts negative feedback to be initially more salient, but to be

dampened down during the subsequent minimization process.

Over the last two decades, the notion of self-enhancement has penetrated the theoretical literature in economics. Imperfect memory has been modeled as largely isomorphic to an inter-personal communication dilemma where disclosing negative information is detrimental while hiding information is costly (Bénabou and Tirole, 2002; Gottlieb, 2014). In consequence, the demand for a positive self-image can optimistically bias the recall process and generate overconfidence (Köszegi, 2006; Gödker et al., 2019). In general, these models predict that negative information on the self tends to be suppressed or biased whenever it is not cognitively too costly, while positive information on the self tends to be prioritized.

This class of models grants large metacognitive control, so that individuals are subject to active self-disinformation. For instance, the first assumption of Bénabou and Tirole (2002) seminal model is that "the individual can, at a cost, increase or decrease the probability of remembering an event or its interpretation." (*ibid.*, p.886).[5] Another peculiar feature of self-enhancement is that it assumes memory to be similar to a storage device, where information is saved or erased. In a nutshell, models of self-enhancing memory claim that people try to maintain a positive self-image through a biased information processing. Hence, people tend to have biased memories with the *goal* of being happy.

Motivational mechanisms have been used to explain biased judgments in a variety of economic situations, from pre-trial bargaining (Babcock et al., 1995) to

[5]This view may seem unrealistic since it concedes individual control over one's memory. However, Bénabou and Tirole (2002) clearly state that they are not assuming people to *directly* suppress memories, but to engage in behaviors that can affect recall, such as intentional rehearsal or avoiding cues. Moreover, intentional control over memories is ordinarily assumed, more than it may seem at first sight. For instance, if one fails to show up at an appointment because she forgets it, she is likely to be blamed for it. Since some degree of memory control is a necessary condition for being morally responsible for memory failures, then memory control is ordinarily granted (see Blustein 2017 for a discussion).

distributional preferences (Deffains et al., 2016), from price setting (Martin et al., 2018) to reciprocal behaviors (Woods and Servátka, 2019). People display both input-control strategies, where some potentially available information is avoided (see Golman et al., 2017, for a review), and internal-biasing strategies, where acquired information is treated in a biased way (see Mele, 1997). Self-enhancing memory belongs to the second category.

Although valence asymmetry is still object of debate among psychologists, in the specific case of information which is relevant for one's self-image, there is consensual evidence on the existence of a positivity bias: favorable information is better recalled than adverse one.[6] In a series of experiments, Sedikides and Green (2000), Sedikides and Green (2004) and Green et al. (2009) observe extensive memory neglect of self-relevant negative information. After a personality test, subjects are presented with an artificial analysis of how much likely they are in engaging in a list of trustworthy and untrustworthy behaviors, kind and unkind behaviors. People tend to recall fewer negative than positive behaviors, even when negative behaviors are consistent with individuals' low self-esteem. Importantly, the phenomenon is directly related to self-image: asymmetric recall disappears as soon as the personality analysis is framed in terms of a third person instead of the self (see also Kuiper and Derry 1982).

A nascent literature in experimental economics has been investigating the question of memory and self-image. It provides evidence of motivationally biased recall for various self-relevant traits, such as trustworthiness and kindness (Li, 2013), generosity (Saucet and Villeval, 2019) and IQ (Zimmermann, 2019; Chew et al., 2018;

[6]This and the following claims refer to non-depressive individuals only. According to the paradigm of depressive realism developed by Mischel (1979) depressed individuals are less subject to self-deceptive behaviors and therefore have more accurate recollection of self-relevant negative feedback. We will not discuss specific traits of depressed individuals, which is beyond the aim of our study.

Li, 2017). Both Zimmermann (2019) and Chew et al. (2018) use a Raven IQ test as a measure of intelligence and investigate how people subsequently recall their performance.[7] In his series of studies on motivated beliefs, Zimmermann (2019) provides participants with a noisy feedback on their relative performance in the Raven test. A few weeks later, he observes that respondents better remember positive than negative feedback. Chew et al. (2018) provide participants with feedback for each Raven matrix and, months later, ask them to recall if they saw the matrix before and if they correctly solved it. Participants not only tend to forget matrices they failed to solve more than matrices they correctly solved, but they are also more likely to misattribute a positive outcome to matrices they have never seen or they did not solve.

2.2.2 Mood-congruent Memory

In psychology, the complex relationship between memory and affect has been modeled in several ways (see Fiedler and Hütter 2013 for a review).[8] Associative theories (Isen et al., 1978; Bower, 1981) explain the mood-congruency phenomenon as being due to the spreading of affective-related items within a memory network. Fiedler (2001) and Bless et al. (2006) treat mood as an associative cue and predicts differential probability of mood-congruency according to the type of cognitive task (accomodation-oriented or assimilation-oriented). Theories of affect-as-information (Schwarz and Clore, 1983; Schwarz, 1988; Schwarz and Clore, 1996) suggest that judgments are marginally adjusted using affect as an available informative signal, thereby causing the potential misattribution of an affective state.

[7] Pioneering evidence on asymmetric recall for intelligence tests can be found in Mischel et al. (1976).

[8] We use the term affect as a generic term, overarching the more fine-grained category of different emotions. We interpret affect (aka affective valence) in a standard way, as measureable on a unidimemsional continuum scale, having negative valence and positive valence at its ends. We adopt the term mood as synonym of affect.

Among economists, Mullainathan (2002) sketches the first economic model of associative memory, where recall outcomes are predicted by the informative cues an individual receives. Nevertheless, mood is never mentioned in his article. Other recent theories on the economic consequences of associative memory can be found in Bordalo et al. (2017) and Bordalo et al. (2019), which explore respectively how events saliency and the representative heuristics distort recalls. Bernheim and Thomadsen (2005) look at the relationship between anticipatory emotions and imperfect memory, but they do not endogeneize memory biases. So far, the only theoretical model of mood-congruency is the recent work by Bodoh-Creed (2017). In line with the psychology literature, Bodoh-Creed models memory as an associative process where the current affective state is a relevant cue for retrieval. In a dynamic setting, he incorporates mood as a deterministic element of biased recall, under the assumption that positive affect increases the probability to recall a positive item rather than a negative one (and vice-versa). By applying the model to financial behavior, he draws predictions on information overreaction and asset price volatility.

The common line of mood-congruency models is to consider recall biases as the byproduct of a heuristic process, by which mood is an associative cue or informative signal. Contrary to self-enhancement, mood-congruency does not need to assume recalls to be intentional and depicts memory to be more similar to a searching device - where information is available or not - rather than to a storage device. Overall, mood-congruent theories claim that people tend to better remember information that is congruent with their current emotional state. Therefore, people tend to have biased positive memories *because* they are happy.

Studies in experimental economics have been investigating the impact of mood on productivity (Oswald et al., 2015), risk-aversion (Nguyen and Noussair, 2014), prudence (Breaban et al., 2016), asset-pricing (Andrade et al., 2015), altruism (Capra, 2004), reciprocity and generosity (Kirchsteiger et al., 2006). Stock prices correlate with exogeneous mood-shifters as diverse as weather (Hirshleifer and Shumway, 2003), seasonal day lengh (Kamstra et al., 2003) and sport events (Edmans et al., 2007).

While economists did not empirically investigate the relationship between affect and memory, psychologists have extensively explored this path. Applied memory literature is replete of examples of mood-congruency: people better retrieve information whose content is congruent with individuals' current affective state (see Blaney (1986) and the literature cited therein).[9] Early evidence can be found in Isen et al. (1978) who show consumers to be significantly happier with a product when they evaluate their past experience in a positive mood. In the specific context of self-relevant feedback recall, Story (1998) asks participants to fill a personality test and provide them with artificial feedback on their traits: she finds that individuals with relatively low self-esteem have a more accurate recall of negative feedback, while individuals with a relatively high self-esteem have a more accurate recall of positive feedback and tend to recall negative feedback as more favorable than it was.[10]

[9]The other fundamental aspect of the interaction between affect and memory is state dependency: people tend to better remember information which was encoded in a similar affective state.

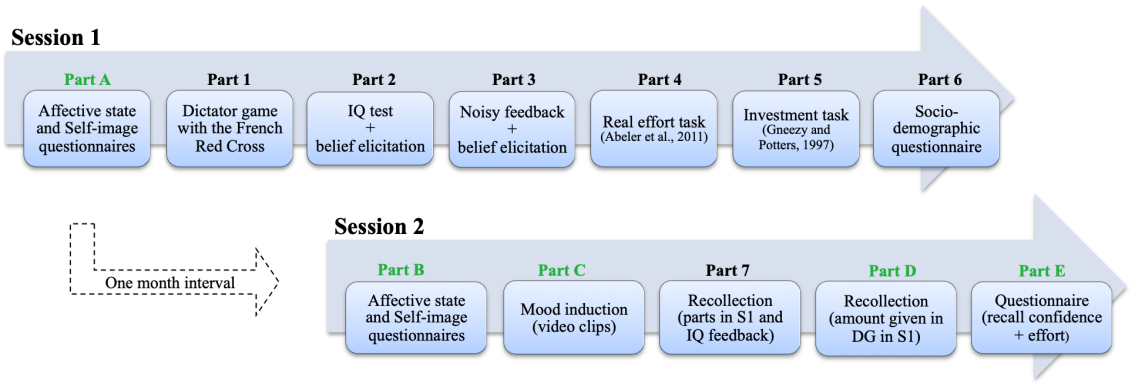
[10]An important empirical question is to distinguish to what extent mood-congruency is a genuine memory phenomenon rather than a general heuristic process. It could be argued that the phenomenon is just a response bias, on the ground that people in positive affective states are more prone to see the world in a good light. In their empirical investigation, Fiedler et al. (2001) reject this option. Their results support the interpretation of mood-congruency as an advantage in retrieval, while they find no evidence of mood-congruency as a superficial response bias.

2.3 Experimental Design and Procedures

We first describe the design of the experiment and then detail the procedures.

2.3.1 Design

To study the existence and relative importance of mood-congruency and self-enhancement on individuals' memory, our design is based on Zimmermann (2019). We set a longitudinal experiment where subjects participated in two sessions, at one-month interval. Parts named by a number (1 to 7) are an accurate replication of Zimmermann (2019)'s design. Parts named by a letter (A to E) are specific to our design. Figure 2.1 displays the timeline of the experiment. We now describe each session in detail.



Note: Parts named by a number (1 to 7, in black) are an accurate replication of the *Recall* treatment in Zimmermann (2019). Parts named by a letter (A to E, in green) are specific to our design.

Figure 2.1: Timeline of the experiment

2.3.1.1 Session 1

Session 1 consisted in seven parts. Parts 1, 4, 5 and 6 were only filler tasks who aimed at obfuscating the purpose of the experiment (see Zimmermann (2019), p.7, for a justification). The experiment started with part A. Subjects were asked to

report their current affective state with respect to different items on a 0-10 Likert scale (UK ONS, 2011) and answered a self-image questionnaire (Schwarzer et al. 1995, see Appendix 1). This part was not payoff-relevant.

In part 1 (filler task), subjects played a one-shot dictator game. They were endowed with 10 euros and could decide if they wanted to donate part of this amount to the French Red Cross.

In part 2, subjects performed an IQ test in the form of 10 Raven matrices to solve in 10 minutes. Before the test, subjects were explicitly told that this kind of questionnaires is frequently used to measure intelligence and performance is correlated with income and education outcomes. After the test, subjects were informed that they had been randomly matched with nine people who previously took the same test.[11] They were ranked within this 10-people group according to their performance in the Raven test. In case of equal score between two or more subjects, the computer randomly broke the tie. Then, we measured subjects' prior beliefs on they relative rank. Individuals were asked to estimate, in percentage, the likelihood that they were in the upper half of the ranked group. They were also asked to estimate, for every possible rank (from 1 to 10), the likelihood they thought it was that they held this rank. Belief elicitation was incentivized through a quadratic score rule, plus a fixed amount of 4 euros.

In part 3, subjects received a noisy feedback about their relative performance. The noisy feedback procedure was based on Eil and Rao (2011): 3 out of the 9 members of the group were randomly selected and the participant was informed, for each of them, if he ranked higher or lower. Thus, each subject could receive

[11]Specifically, the comparison group is made of 64 subjects with similar demographic characteristics and from the same geographical area, who took the same test at GATE-Lab in 2015.

four kinds of feedback: 3 people performed better than him, 2 people performed better than him, 1 person performed better than him or 0 person performed better than him. To rule out inattention, subjects were asked to repeat the information right after receiving the feedback. Then, we measured subjects' posterior beliefs on their relative rank. Individuals were asked to estimate, in percentage, the likelihood that they were in the upper half of the ranked group.

Part 4 (filler task) consisted in a real-effort task, similar to Abeler et al. (2011): subjects had five minutes to count the number of zeros in tables containing zeros and ones and had to report the correct number for each table. They were paid 0.2 euro per correct report, plus a fixed amount of 5 euros.

In part 5 (filler task), subjects were endowed with 2 euros and had to decide how many cents to put in a risky investment in which the amount invested has one chance out of three to be multiplied by 2.5 and two chances out of three to be void (Gneezy and Potters, 1997). They received a fixed payment of 3 euros in addition to the investment return.

Finally, in part 6 (filler task) subjects filled a socio-demographic questionnaire, which paid a fixed amount of 5 euros.

2.3.1.2 Session 2

Session 2 took place one month later and consisted in five parts. Session 2 combines the recall elicitation task of Zimmermann (2019) with a mood induction procedure based on Andrade et al. (2015).

Session 2 started with part B. Subjects were asked to report their affective state and to answer a self-image questionnaire. This part mirrored part A in session 1 and aimed at controlling for the idiosyncratic baseline levels and variations, in both affect and self-esteem.

Part C was treatment-specific. Subjects watched a video clip which combined two excerpts from commercial movies. Depending on the treatments group, the clip was meant to induce either positive, negative or neutral affective state. In the Positive treatment, people watched an excerpt from "The Dinner Game" and one from "Les trois frères".[12] In the Negative treatment, the excerpts were from "American History X" and "Schindler's List". In the neutral treatment, they were from "Blue" and "The Lover". Video clips were carefully chosen from the database of Schaefer et al. (2010), that assesses the relative efficiency of a large sample of clips in inducing different emotions.[13] For each treatment, we selected the two excerpts that have been ranked as the most effective ones in inducing – respectively - amusement and negative affect, as well as two neutral excerpts for the control group. After watching the video, subjects were asked to report their affective reaction using a Self-Assessment Manikin (Bradley and Lang, 1994), where they assessed the affective valence (from "clearly positive" to "clearly negative") and intensity (from "not intense at all" to "very intense") of their emotional experience. In addition, they were asked to choose from a list of emotions the one that best described their feeling. These self-reported measures, which are very common in the literature on emotions, were meant to check that the treatments triggered the

[12]These films are famous comedies in France. According to www.senscritique.com, "The Dinner Game" in ranked the 2nd best French comedy of all times and had more than 9 million admissions in the box office. "Les trois frères" is ranked the 10th best French comedy and had more than 6.8 million admissions in the box office.

[13]Schaefer et al. (2010)'s study was conducted less than ten years ago, using videos in French language, on a sample of French-speaking European students, which make it particularly suitable for replication in our context. In their study, 364 subjects viewed the video clips in individual laboratory sessions and rated each video on multiple dimensions. Results shows that the video clips were effective with regard to several criteria such as positive and negative affect, arousal and emotional discreteness.

desired mood. Both our mood-induction technique and our mood-manipulation check closely replicated Andrade et al. (2015)'s design.

Part 7 dealt with two recall tasks. Subjects were asked to briefly summarize each part of session 1 in one sentence and were paid 0.5 euro for a sufficiently accurate description. Once the summary was completed, they were recalled that in session 1 they took an IQ test and received feedback about their relative performance within a randomly selected group. Thereafter, we asked them to recall and report how many people ranked higher than them. Possible answers were "3", "2", "1", "0" or "I don't recall". Subjects were paid 2 euros if they recalled correctly. This memory elicitation procedure accurately followed the design of the *Recall* treatment in Zimmermann (2019).

In part D, we collected one additional piece of information. Subjects were asked to recall the amount of their donation to the French Red Cross charity in part 1, session 1. This measure allowed us to glimpse potential interactions between recalls of two self-relevant traits: intelligence (feedback on IQ) and generosity (donation to a charity).

Finally, in part E, subjects answered a short questionnaire providing some feedback on the two sessions. In particular, they reported on a 10-item scale how much confident they were with each recall task and how much effort they provided to retrieve the information at stake.

2.3.2 Procedures

The experiment was programmed using Z-tree (Fischbacher, 2007). It was conducted at GATE-Lab (Lyon, France). The first session lasted on average 50 min-

utes. The second session lasted on average 40 minutes. To reduce attrition and to incentivize subjects to show-up in the second session, all payments from the experiment were made at the end of the second session. At the end of the first session, subjects only received 5 euros payoff for their participation. Moreover, they were given a slip of paper stating the exact date and time of the second session and were reminded twice via email about the second lab session. One month after, subjects came back to the laboratory on the same weekday and daytime as in session 1. They received 15 euros payoff plus the earnings made in one part of the experiment randomly selected for payment. Subjects were informed of this payment scheme at the very beginning of the experiment.

A total of 250 subjects were recruited, using Hroot (Bock et al., 2014). 25 subjects did not participate to the second session and are therefore excluded from the analysis.[14] 96 subjects participated in the Neutral treatment, 60 in the Positive treatment and 69 in the Negative treatment.[15] Table A2.1 in Appendix 2 summarizes the subjects' characteristics in each treatment.

2.4 Behavioral Conjectures

The following section formulates the hypotheses predicted by the two competing theories: self-enhancing memory and mood-congruent memory. We define positive and negative feedback as follows: feedback is positive if the subject has received

[14] In Zimmermann (2019), subjects that did not show up for the second lab session received an email with a Qualtrics link that allowed them to complete the study online within the following 24 hours. This was not feasible in our case since we needed subjects to watch video clips in a controlled environment.

[15] More data will be collected to reach 120 subjects per treatment. This is the required number to replicate the data analysis in Zimmermann (2019) (N=118 in the *Recall* treatment). The temporarily imbalanced sample size observed between treatments results from a deliberate choice to collect data in the Neutral treatment as a priority since it aims at replicating Zimmermann (2019)'s study and is the baseline of our study.

at least two positive comparisons out of three comparisons. Reversely, feedback is negative if the subject has received zero or one positive comparison out of three comparisons. A recall is correct if the participant recalls exactly the number of positive comparisons he received in session 1.

In our design, each subject can be in one out of six possible states of the world (2×3): they received either positive or negative feedback in session 1, and they are assigned either to the Neutral, Positive or Negative treatment in session 2. These six states of the world are represented by the six cells in Table 2.1. Each letter describes the percentage of subjects who correctly recalled their number of positive comparisons, in each possible state of the world. For instance, letter A should be interpreted as the percentage of people who had a correct recall among the people in the Neutral treatment who received a positive feedback. Since we use a between-subject design, each case is computed on different individuals. Table 2.2 summarizes the predictions of both self-enhancement and mood-congruency effects.

Table 2.1: Percentage of correct recalls conditional on the valence of feedback and of the mood.

	Participants' mood		
	Neutral	Positive	Negative
Positive feedback	**A**	**C**	**E**
Negative feedback	**B**	**D**	**F**

Note: Negative feedback =1 if at least 2 out of the 3 comparisons with the randomly selected group members are negative.

Example: Among the people in the positive treatment who received a positive feedback, $C\%$ recall their number of positive comparisons correctly.

Table 2.2: Predictions

Zimmermann (2019) predicts	Self-enhancing memory predicts	Mood-congruent memory predicts	Interpretation
$\boldsymbol{A} > \boldsymbol{B}$	$\boldsymbol{A} > \boldsymbol{B}$ $\boldsymbol{C} > \boldsymbol{D}$	$\boldsymbol{A} > \boldsymbol{B}$ $\boldsymbol{C} > \boldsymbol{D}$	Consistent with self-enhancing *and* mood-congruent memory
		$\boldsymbol{C} > \boldsymbol{E}$ $\boldsymbol{D} < \boldsymbol{F}$	Consistent with mood-congruent memory
	$\boldsymbol{E} > \boldsymbol{F}$	$\boldsymbol{E} < \boldsymbol{F}$	Tests the relative dominance of each effect

Note: Each letter describes the percentage of subjects who had a correct recall in each possible state of the world in Table 2.1.

The Neutral treatment aims at replicating Zimmermann (2019) who finds that individuals recall negative feedback with less accuracy compared to positive feedback. Our first conjecture is a replication of Zimmermann (2019).

Conjecture 1 (Replication of Zimmermann, 2019) *: Under neutral mood*[16]*, subjects recall negative feedback with less accuracy than positive feedback. Formally,* $\boldsymbol{A} > \boldsymbol{B}$.

In the Positive treatment (as in the Neutral treatment), both self-enhancement and mood-congruency predict subjects to better recall positive feedback, although for different reasons. Self-enhancing memory predicts positive feedback to meet an individual's demand for self-esteem. When the content of the feedback is potentially harmful for oneself, subjects have motivational reasons to forget it. Mood-congruency predicts people in non-negative mood to better recall non-negative feedback. Since most people tend to be in a non-negative affective state –in particular, youngsters (Eurostat, EU-SILC 2013)– mood congruency predicts positive feedback to be, on average, more accessible to recall. Therefore, observing higher percentage of correct recall for positive than for negative feedback in the Neutral

[16]Neutral mood means that participants' mood was not manipulated. Therefore, it can also be understood as "in the absence of mood manipulation".

and Positive treatment is consistent both with self-enhancement effects and mood-congruency effects.

To identify the existence of mood-congruency alone, one needs to compare the percentage of correct recall between treatments. Mood-congruent memory predicts that being in a negative mood will trigger recall of negative feedback and, reversely, that being in a positive mood will trigger recall of positive feedback. Therefore, if mood-congruency plays a role in memory retrieval, negative feedback should be recalled more accurately in the Negative treatment than in the Positive treatment, and vice-versa positive feedback. Self-enhancement is silent about between-treatment comparisons. We state our second conjecture as follows:

Conjecture 2 (Mood-congruent Memory) *: Subjects recall negative feedback with more accuracy under negative mood than under positive mood. Symmetrically, subjects recall positive feedback with more accuracy under positive mood than under negative mood. Formally,* $\boldsymbol{C} > \boldsymbol{E}$ *and* $\boldsymbol{D} < \boldsymbol{F}$.

Self-enhancing and mood congruent memories predict similar outcomes in most *but not all* situations. In the Negative treatment, self-enhancement predicts subjects to better recall positive feedback. In contrast, mood-congruency predicts individuals to better recall feedback that is congruent with their mood, i.e., negative feedback. Therefore, the particular case in which participants are in a negative mood allows us to disentangle relative dominance of the two effects. If self-enhancing memory dominates, people should better recall feedback if the latter is positive, regardless of their current mood. Therefore, self-enhancing memory predicts $E > F$. If mood-congruent memory dominates, people should better recall feedback if the latter is negative, since negative feedback is more accessible than positive one when mood is negative. Hence, mood-congruency predicts

$E < F$. Importantly, neither outcome discards the existence of the alternative effect, however it discards its dominance. We state our third conjecture as follows:

Conjecture 3 (Relative dominance of self-enhancing *vs.* mood-congruent memory)

- *If self-enhancing memory dominates, subjects under negative mood recall more accurately positive feedback than negative feedback. Formally,* $\boldsymbol{E > F}$.
- *If mood-congruent memory dominates, subjects under negative mood recall more accurately negative feedback than positive feedback. Formally,* $\boldsymbol{E < F}$.

2.5 Results

A prerequisite of our experiment is that mood induction in Part C was effective. Therefore, before presenting the results about memory effects, we first carefully check whether Part C produced the desired affective states and report results regarding participants' mood elicitation.

2.5.1 Mood Induction

Participants' mood was elicited through two channels traditionally used to investigate individuals' emotions: emotional valence and emotional arousal.

Results on Emotional Valence: Emotional valence corresponds to the value associated with a stimulus, as expressed on a continuum from pleasant to unpleasant. In our design, the stimulus was a video clip watched at the beginning of session 2. After watching the video clip, a high self-reported valence indicates that the stimulus was rather pleasant while a low self-reported valence indicates

that the stimulus was rather unpleasant. Figure 2.2 shows the distribution of self-reported valence and arousal by treatment, on a two-dimension Valence-Arousal space. Each dot represents one observation. It can be visually inferred that in the Positive treatment (green squares) and Negative treatment (red triangles) participants experienced a very different emotional valence (horizontal axis), compared to participants in the Neutral treatment (black circles). Table 2.3 reports the average reported emotional valence and arousal *after* the treatment manipulation, by treatment. The average reported valence (V hereafter) in the Positive treatment (V=7.97) is significantly higher than the average reported valence in the Neutral treatment (V=5.36, $p<0.001$, Mann-Whitney).[17] Similarly, the average reported valence in the Negative treatment (V=1.80) is significantly lower than the average reported valence in the Neutral treatment ($p<0.001$, MW). The distribution of the reported valence is also significantly different between treatments ($p<0.001$, Kolmogorov-Smirnov tests for all pairwise comparisons). Therefore, participants in the Positive treatment had a significantly (at a 1% level) more pleasant experience than participants in the Neutral and Negative treatments, and participants in the Negative treatment had a significantly (at a 1% level) less pleasant experience than participants in the Neutral and Positive treatments.

Results on Emotional Arousal: Emotional arousal corresponds to the self-reported intensity associated with a stimulus, as expressed on a continuum from not intense at all to very intense (9-item scale). It can be inferred from Figure 2.2 that arousal (vertical axis) is higher in the Positive and Negative treatments than in the Neutral treatment. Table 2.3 shows that emotional arousal (A hereafter) is significantly higher in the Negative (A=7.17) and Positive (A=4.97) treatments than in the Neutral treatment (A=3.31, $p<0.001$, MW tests). The distribution

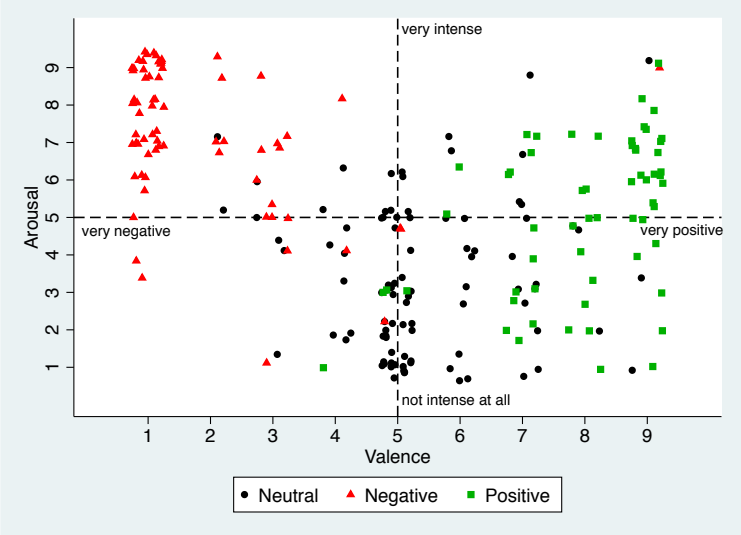
[17] In all non-parametric tests reported in this paper, each individual gives one independent observation, and all tests are two-sided.

of reported arousal is also significantly different between treatments ($p<0.001$, Kolmogorov-Smirnov tests). Therefore, participants in the Positive and Negative treatments experienced significantly (at a 1% level) more intense emotions than participants in the Neutral treatment.

Table 2.3: Summary statistics - Participants' mood *after* treatment manipulation

	Treatment		
	Neu.	Pos.	Neg.
	(1)	(2)	(3)
Emotional Valence	5.36	7.97***	1.80***
(1: clearly negative; 9: clearly positive)	(1.35)	(1.25)	(1.43)
Emotional Arousal	3.31	4.97***	7.17***
(1: not intense at all; 9: very intense)	(1.95)	(2.03)	(1.86)
N	96	60	69

Notes: p-values are from two-tailed Mann-Whitney tests. The Positive and Negative treatments are compared to the Neutral treatment. Standard deviation in parentheses. One observation per individual. *** $p<0.01$.



Note: each dot represents one individual. For a better view, we used the "jitter" option in Stata that differentiates dots located in the same position.

Figure 2.2: Valence-Arousal 2-dimensional space, by treatment

Participants were also asked to choose from a list of six sets of emotions the one that best described their emotion after having watched the video. Participants

could choose only one category of emotions. If watching the video successfully triggered different affective states, participants should have reported different categories of emotions depending on the treatment they were assigned to. Table 2.4 reports the percentage of participants by category of emotions and by treatment. In the Positive treatment, 90% of the participants reported that the category of emotions that best described their feeling was either "Excited, enthusiastic, happy" or "Calm, relax, peaceful", which in both cases are positive emotions. By contrast, in the Negative treatment 89.85% of the participants stated that they were either "Anxious, scared, terrified" or "Sad, depressed, unhappy", which are both negative emotions. In the Neutral treatment, 57.29% of the participants chose a category with a neutral valence (either "Bored, disinterested, jaded" or "Neutral (no emotional reaction)").

Table 2.4: Reported category of emotions that best described actual affective state *after* treatment manipulation (in %)

	Treatment		
	Neu.	Pos.	Neg.
Anxious, scared, terrified	10.42%	0%	55.07%
Sad, depressed, unhappy	4.17%	0%	34.78%
Bored, disinterested, jaded	5.21%	3.33%	2.90%
Neutral (no emotional reaction)	52.08%	6.67%	5.80%
Excited, enthusiastic, happy	2.08%	68.33%	1.45%
Calm, relax, peaceful	26.04%	21.67%	0%
N	96	60	69

Imported Emotions: Since our treatments manipulate participants' mood, it is important to control for participants' mood when they arrived at the lab. Indeed, it is possible that participants' mood is affected by characteristics independent from the experiment (e.g., weather, events experienced before coming to the experiment, etc.), and such potential heterogeneity needs to be controlled both within and between treatments. At the beginning of the first and the second session, participants were asked to fill in questionnaires related to their affective state

and to their self-image. These non-incentivized questionnaires allowed us to control for idiosyncratic baseline levels as well as heterogeneous exogenous variations, both in affective state and self-esteem. Table A2.2 in Appendix 2 summarizes the average answers in this questionnaires, both across sessions and across treatments. It shows that participants' affective state and self-esteem when they arrived at the lab are not significantly different either between treatments (Mann Whitney-tests), or between sessions (Wilcoxon signed-rank test). Therefore, the different moods observed between treatments result from our treatment manipulation and not from idiosyncratic heterogeneity or variations between the first and second sessions initial states.

Overall, these results show that the treatment manipulation (watching different video clips) successfully induced the desired affective states.

2.5.2 Memory Accuracy

First, we investigate the average percentage of correct recalls of feedback in the IQ test. A recall is correct if the participant recalls exactly the number of positive comparisons he received in session 1. Then, we consider an alternative measure of recall. Instead of investigating the percentage of correct recalls of feedback, we investigate how well subjects recall the parts related to the IQ test.[18]

2.5.2.1 Memory Accuracy of feedback

Figure 2.3 depicts the percentage of correct recalls of feedback for the different levels of comparisons, by treatment. Results from the *Recall* treatment in Zimmermann (2019), which exactly corresponds to the parts named by a digit in Figure 2.1, are represented by the black-dashed line. Our Neutral treatment (black line)

[18]This alternative measure of recall is also considered by Zimmermann (2019), Section 3.2.3 p.19.

is expected to replicate and replicates results from Zimmermann (2019) when participants have received 1, 2 or 3 positive comparisons: the percentage of correct recalls substantially increases with the number of positive comparisons.[19]

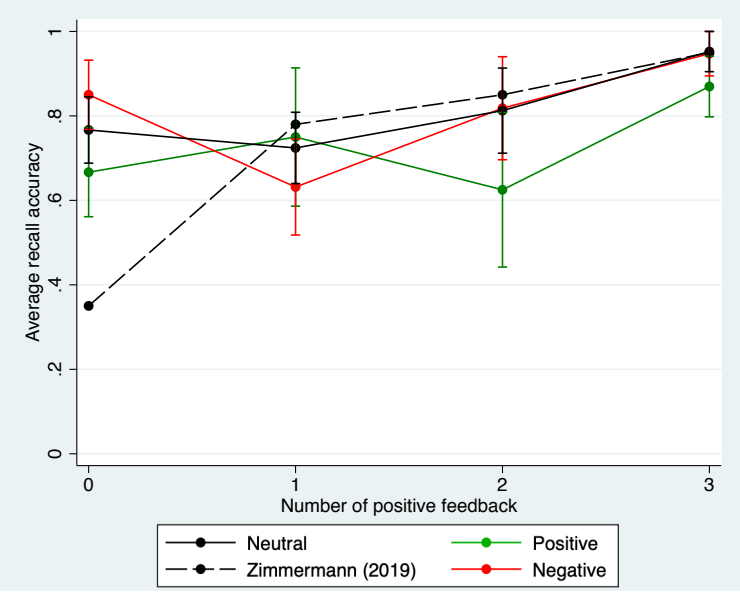


Figure 2.3: Average recall accuracy for different levels of comparisons, by treatment

We introduce our first result:

Result 1 (Replication of Zimmermann, 2019) : *In the Neutral treatment, subjects recall negative feedback with less accuracy than positive feedback.*

Result 1 supports Conjecture 1.

Support for Result 1: Table 2.5 displays subjects' percentage of correct recalls, conditional on receiving positive feedback (2 or 3 positive comparisons) or negative

[19]When participants have received zero positive comparisons, we need to know the number of observations and the standard deviation in Zimmermann (2019)'s data to test whether the means are significantly different between Zimmermann (2019)'s data and our Neutral treatment.

feedback (0 or 1 positive comparisons), by treatment. In the Neutral treatment (column 1) that replicates the *Recall* treatment in Zimmermann (2019), the average percentage of correct recalls is 89% when the feedback is positive and 75% when the feedback is negative (Table 2.5). The difference is significantly different at a 10% level (p=0.082, MW, N=96).[20]

Table 2.5: Average recall accuracy

	Treatment			
Treatment	Neu.	Pos.	Neg.	All
Pos. feedback (1)	0.89	0.81	0.90	0.87
	(0.31)	(0.40)	(0.31)	(0.34)
Neg. feedback (2)	0.75	0.69	0.74	0.73
	(0.44)	(0.47)	(0.44)	(0.44)
p-value (1)-(2)	*0.082*	*0.301*	*0.102*	*0.014*
N	96	60	69	225

Note: Negative feedback =1 if at least 2 out of the 3 comparisons with the randomly selected group members are negative. P-values are from Mann-Whitney tests. The average recall accuracy is not significantly different between treatments, regardless of the valence of the feedback. One observation per individual. Standard deviation in parentheses.

Example: Among the subjects in the Neutral treatment who received positive feedback, 89% recall their feedback correctly.

Table 2.6 provides coefficients from linear probability models in which the dependent variable is a dummy that is 1 if the participant correctly recalled the number of positive comparisons and 0 otherwise. In Model (1) the independent variables are a dummy that is equal to 1 if feedback was negative and 0 if it was positive, and the three treatments (with the Neutral treatment as the reference category). Model (2) includes the interactions between the valence of feedback (positive *vs.* negative) and the treatments (Neutral, Positive and Negative). Models (3) and (4) additionally control for the predicted belief adjustment defined as the belief adjustment if subjects would follow Bayes' rule, and the Rank which refers to subject's rank in their group. These control variable allows us to replicate

[20] Assuming a power level of 0.8, the required sample size would be 116 observations per group to have a difference significant at a 5% level.

Zimmermann (2019), Table 3 (p.15). Model (1) shows that subjects that obtained negative feedback recall this feedback with significantly less (at a 1% level) accuracy one month later, compared to subjects that received positive feedback. This is also the case (at a 10% level) when controlling for the interaction terms between the feedback and the treatments (Models (2) and (3)), and the predicted belief adjustment (Model (3)).

Overall, these results replicate the findings of Zimmermann (2019): subjects recall negative feedback with lower accuracy compared to positive feedback. One main explanation for these results is that subjects' memory is retrieved self-servingly to enhance individuals' self-image. This self-enhancement effect predicts that individuals better recall positive than negative feedback to enhance their self-image, which is consistent with the results in the Neutral treatment. Another possible explanation, however, is that individuals simply retrieve feedback that is associated to their current affective state. This associative recall mechanism could explain why most individuals (those in positive or non-negative affective state) would naturally better recall positive feedback, and could be at stake in Zimmermann (2019)'s experiment in which participant's mood is neither controlled nor varied.[21] We now investigate whether the observed participants' asymmetric recall on feedback could be driven by mood congruency.

We now introduce our second result:

[21]If we consider that experimental subjects are usually undergraduate students, their affective state is very likely to be on the positive spectrum. More than 7 out of 10 Europeans aged between 16 and 24 report to have been happy all or most of the time during the four weeks prior to the survey. Less than one over 10 reports to have been rarely or not happy at all (Eurostat, EU-SILC 2013). Indeed, one of the most robust findings in happiness research is that well-being is U-shaped over the life cycle, so that young adults tend to be happier than average (Blanchflower and Oswald, 2008; Van Landeghem, 2012). This is true for cognitive measures (life satisfaction) but also for affective measures (mood).

Table 2.6: Recall Accuracy

	Recall Accuracy			
	(1)	(2)	(3)	(4)
=1 if neg. feedback	-0.141***	-0.146*	-0.153*	-0.024
	(0.052)	(0.077)	(0.083)	(0.118)
Neutral treatment	*ref.*	*ref.*	*ref.*	*ref.*
	-	-	-	-
Positive treatment	-0.071	-0.085	-0.084	-0.103
	(0.069)	(0.089)	(0.089)	(0.084)
Negative treatment	0.003	0.008	0.010	0.004
	(0.062)	(0.076)	(0.076)	(0.072)
=1 if neg. feedback * Neu. treat.		*ref.*	*ref.*	*ref.*
		-	-	-
=1 if neg. feedback * Pos. treat.		0.029	0.027	0.074
		(0.137)	(0.137)	(0.134)
=1 if neg. feedback * Neg. treat.		-0.010	-0.010	0.002
		(0.119)	(0.119)	(0.119)
Predicted belief adjustment			0.001	0.001
			(0.002)	(0.001)
rank				-0.031*
				(0.017)
constant	0.889***	0.892***	0.883***	0.976***
	(0.047)	(0.052)	(0.052)	(0.059)
N	225	225	225	225

Note: Results are from a linear probability model of the likelihood to correctly recall the feedback. Negative feedback =1 if at least 2 out of the 3 comparisons with the randomly selected group members are negative. Robust standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Result 2 (Existence of mood-congruent memory) : *Subjects do not recall more accurately feedback that is congruent with their mood. More precisely, subjects in the Negative treatment do not recall more accurately negative feedback than subjects in the Positive treatment, and subjects in the Positive treatment do not recall more accurately positive feedback than subjects in the Negative treatment.*

Result 2 rejects Conjecture 2.

Support for Result 2: The mood-congruency effect predicts that individuals recall with higher accuracy feedback that is congruent with their mood. There-

fore, if mood-congruency affects participants' memory accuracy on feedback, one should observe higher percentage of correct recalls when the valence of the mood is of same nature than the valence of the feedback to be recalled. There are two situations in which feedback and mood are of same valence: when subjects had to recall positive feedback in the Positive treatment and when subjects had to recall negative feedback in the Negative treatment. If mood-congruency plays a role in memory retrieval, one should observe higher percentage of correct recalls in these two situations than when feedback and mood are of different valence. When participants have to recall positive feedback and are in a positive mood (Positive treatment), the percentage of correct recalls is 81% (see Table 2.5). When participants have to recall positive feedback but are in a negative mood (Negative treatment), the percentage of correct recalls is 90% (p=0.307, MW). Therefore, after receiving positive feedback, the percentage of correct recalls is not higher when feedback and mood are of same valence than when they have opposite valence. When participants have to recall negative feedback and are in a negative mood (Negative treatment), the percentage of correct recalls is 74%. When participants have to recall negative feedback but are in a positive mood (Positive treatment), the percentage of correct recalls 69%. This could be evidence of mood-congruence. However, the difference is not significant (p=0.623, MW).

Looking at the interaction terms in Table 2.6 allows us to investigate further whether mood-congruency may play a role in feedback retrieval. If, as predicted by mood-congruency, negative feedback are better recalled when subjects are in a negative mood, the interaction term between the valence of the feedback (=1 if negative) and the Negative treatment should be positive and significant, meaning that being in the Negative treatment increases the likelihood of recalling correctly one's feedback, compared to when being in the Neutral treatment. This is not the

case (p=0.909, Table 2.6).

Overall, these results show no clear evidence of mood-congruent memory. Individuals do not exhibit higher percentage of correct recalls when the feedback to retrieve is congruent with their mood.

We now introduce our third result:

Result 3 (Relative Dominance) : *In the Negative treatment, subjects recall more accurately positive feedback than negative feedback. This result is consistent with self-enhancing memory but not with mood-congruent memory.*

Result 3 supports the dominance of self-enhancing memory.

Support for Result 3: Table 2.5 shows that, in the Negative treatment, the percentage of correct recalls is 90% when feedback is positive and 74% when feedback is negative. The difference is significant at the 90% confidence threshold (p=0.102, N=69, MW).[22] Moreover, the negative coefficient of the interaction term between the negative feedback and the Negative treatment in Table 2.6 shows that participating in the Negative treatment does not alleviate the asymmetric recall bias observed after subjects received positive *vs.* negative feedback: subjects that obtained negative feedback recall this feedback with significantly less accuracy one month later, even in the Negative treatment.

[22]In the Negative treatment, the required sample size would be 71 observations per group for a difference significant at a 10% level, and 90 observations per group for a difference significant at a 5% level.

2.5.2.2 Memory Accuracy on IQ-related parts

To investigate further the driving force of asymmetric feedback recall, we consider an alternative measure of recall and explore the recollection of IQ-related *vs.* non IQ-related parts of the experiment. If memory accuracy is consistent with the self-enhancement effect, one should observe lower average recall accuracy of IQ-related parts for subjects that received negative feedback, compared to subjects who received positive feedback, regardless of the treatment. If mood-congruent memory also plays a role, one should observe higher average recall accuracy of IQ-related parts for subjects that received negative feedback in the Negative treatment than for subjects who received negative feedback in the Positive treatment. Symmetrically, one should observe higher average recall accuracy of IQ-related parts for subjects that received positive feedback in the Positive treatment than for subjects who received positive feedback in the Negative treatment.

Table 2.7 reports the average number of IQ-related parts correctly recalled by subjects. On average, subjects who received positive feedback recall correctly 1.46 IQ-related parts. Subjects who received negative feedback recall on average 1.09 IQ-related parts. The difference is highly significant ($p<0.001$, MW). This tendency is also observed when taking each treatment separately (see Table 2.7). By contrast, the average number of recalled non IQ-related parts is not higher when subjects received positive rather than negative feedback (see Table 2.7), thus confirming memory to be selective with respect to the IQ parts only.

Regarding the existence of mood-congruency, Table 2.7 shows that subjects who received positive feedback in the Positive treatment recall on average 1.68 IQ parts. Subjects who received positive feedback in the Negative treatment recall on average 1.43 parts. This is consistent with mood-congruency (more accurate

recall when mood is congruent with feedback), but the difference is not significant (p=0.262, MW).[23] Subjects who received negative feedback in the Negative treatment recall on average 1.15 IQ-related parts. Subjects who received negative feedback in the Positive treatment recall on average 1.03 IQ-related parts. This is also consistent with mood-congruency, but again, the difference is not significant (p=0.544, MW).[24]

To explore the relative dominance of each effect, we can compare the average number of IQ-related parts correctly recalled in the Negative treatment, depending on the nature of the feedback. When subjects received positive feedback, they recalled on average 1.43 IQ-related parts. When subjects received negative feedback, they recalled on average 1.15 IQ-related parts (p=0.089, MW). This finding supports the relative dominance of the self-enhancement effect over mood-congruency, with 91% confidence.

Table 2.7: Average number of recalled parts (IQ *vs.* Non IQ related)

	Recall Accu. of IQ parts				Recall Accu. of Non IQ parts			
Treatment	Neu.	Pos.	Neg.	All	Neu.	Pos.	Neg.	All
Pos. info (1)	1.30	1.68	1.43	1.46	1.32	1.32	1.53	1.38
	(0.81)	(0.60)	(0.82)	(0.76)	(1.29)	(1.05)	(1.14)	(1.16)
Neg. info (2)	1.08	1.03	1.15	1.09	1.46	1.55	1.28	1.43
	(0.82)	(0.87)	(0.74)	(0.80)	(1.13)	(1.12)	(1.17)	(1.14)
p-value (1)-(2)	*0.202*	*0.002*	*0.089*	*<0.001*	*0.400*	*0.471*	*0.348*	*0.742*
N	96	60	69	225	96	60	69	225

Note: Negative feedback =1 if at least 2 out of the 3 comparisons with the randomly selected group members are negative. P-values are from Mann-Whitney tests. None of the average recall accuracy are significantly different treatment, except the average recall accuracy when the feedback to recall was positive, between the Neutral and Positive treatment (p=0.039, Mann Whitney test). One observation per individual. Standard deviation in parentheses.

[23] Assuming a power level of 0.8, the required sample size would be 102 observations per group to have a difference significant at a 10% level, and 130 observations per group for a type-I error rate of 0.05.

[24] Assuming a power level of 0.8, one would require more than 550 observations per group to have a difference significant at a 10% level, and more than 700 observations per group for a difference significant at a 5% level.

Overall, these results –especially those observed in situations in which self-enhancement and mood-congruency predict different memory outcomes– provide support for the existence of the former effect over the later. This is observed regardless of the variable used to measure recall accuracy: memory of feedback itself or memory of the parts related to the IQ test.

In the following section, we investigate subjects' memory errors, which refer to the direction and magnitude of subjects' recalls when they do *not* recall their feedback correctly. Investigating memory errors is interesting for two reasons. First, it allows us to investigate whether memory errors are symmetrically distributed around zero (imperfect memory), or whether they are biased (motivated memory). Second, it allows us to explore another dimension of subjects' memory in which both self-enhancement and mood-congruency may play a different role than the one observed when focusing on memory accuracy only.

2.5.3 Direction of Memory Errors

We define a recall error as the difference between the actual number of negative comparisons and the recalled number of negative comparisons (Zimmermann, 2019). A positive value means that the subject has underestimated the number of negative comparisons and thus exhibits an optimistic recall bias. By contrast, a negative value means that the subject has overestimated the number of negative comparisons and thus exhibit a pessimistic recall bias.

Table 2.8 summarizes the results from a linear probability model in which the dependent variable is the recall error. In Model (1) the independent variables include a dummy equal to 1 if the feedback to recall was negative and 0 if it was positive, and the three treatments (with Neutral treatment as the reference cat-

egory). Model (2) includes also interaction terms between the feedback (positive *vs.* negative) and the treatments (Neutral, Positive and Negative). Models (3) and (4) additionally control for the predicted belief adjustment defined as the belief adjustment if subjects would follow Bayes' rule, and the Rank which refers to subject's rank in their group. Table 2.8 shows that subjects who received negative feedback misremember in an optimistic way. Namely, they overestimate the number of positive comparisons they received. This is the case regardless of the model specification.

The coefficient of the third interaction term in Table 2.8 (which is equal to 1 if subjects had to recall negative feedback in the Negative treatment) is negative and weakly significant, at a 10% level. This may suggest instances of mood congruency, since subjects who received negative feedback in the Negative treatment exhibit less optimistic recall than subjects who received negative feedback in the Neutral treatment. Regardless, self-enhancement effects still dominate: overall, the model predicts memory errors to be positive in each treatment.

Overall, these results show that, just like memory accuracy, memory errors are self-serving: when they do not recall correctly, individuals exhibit overly optimistic recalls of past feedback. By contrast, we do not find clear evidence of mood-congruency, although the weakly significant (at a 10% level) negative coefficient estimated for participants in the Negative treatment may suggest a mitigating effect of negative mood on optimistic memory errors.

Table 2.8: Memory errors

	Memory errors			
	(1)	(2)	(3)	(4)
=1 if neg. feedback	0.316***	0.366***	0.361***	0.457***
	(0.070)	(0.117)	(0.118)	(0.157)
Neutral treatment	*ref.*	*ref.*	*ref.*	*ref.*
	-	-	-	-
Positive treatment	0.029	-0.031	-0.031	-0.044
	(0.096)	(0.119)	(0.119)	(0.115)
Negative treatment	-0.032	0.129	0.130	0.126
	(0.078)	(0.101)	(0.101)	(0.099)
=1 if neg. feedback * Neu. treat.		*ref.*	*ref.*	*ref.*
		-	-	-
=1 if neg. feedback * Pos. treat.		0.138	0.137	0.172
		(0.190)	(0.192)	(0.196)
=1 if neg. feedback * Neg. treat.		-0.281*	-0.281*	-0.272*
		(0.150)	(0.150)	(0.151)
Predicted belief adjustment			0.0004	0.001
			(0.002)	(0.002)
Rank				-0.023
				(0.022)
Cons.	-0.133***	-0.162*	-0.168**	-0.099
	(0.045)	(0.082)	(0.084)	(0.088)
N	225	225	225	225

Note: Linear probability model of the difference between recalled and actual number of positive feedback. Negative feedback =1 if at least 2 out of the 3 comparisons with the randomly selected group members are negative. Rank refers to subject's rank in their group. Predicted belief adjustment is defined as the belief adjustment if subjects would follow Bayes' rule. Standard errors in parentheses. One observation per individual. * p<0.10, ** p<0.05, *** p<0.01.

2.6 Conclusion

This study is the first experiment aimed at jointly testing two well-documented memory effects: mood-congruency and self-enhancement. Our design creates controlled situations where the two theories predict different outcomes and thereby allows us to identify their respective roles in the dynamic of false beliefs about the self. We provide empirical ground for the ongoing debate in bounded rationality research about how to interpret rationality failures: on the one hand, the self-enhancement models grant large meta-cognitive control; on the other hand, congruency models assume an underlying heuristic process without intentionality.

In doing so, we mediate between two parallel strands of literature.

Our laboratory experiment combines the designs by Zimmermann (2019) and Andrade et al. (2015) to study self-relevant feedback recollection under different moods. Our results replicate and extend Zimmermann (2019)'s. First, everything else equal, people tend to display better recall of positive than negative feedback. Second, individuals exhibit overly optimistic recalls: they overestimate the number of positive feedback and under-estimate the number of negative feedback. On the contrary, even though our mood-manipulation proves to be effective in inducing the desired affective state, we find no or limited evidence of the effect of mood on asymmetric recall. Individuals do not recall more accurately feedback that is congruent to their mood.

One possible explanation for the null results of mood-congruency is the size of our experimental sample. If the detectable effect of mood-congruency is smaller than the one of self-enhancement, we may fail to observe the former because of a lack of power in our tests. Importantly, this explanation would corroborate the dominance of self-enhancement effects. Although our standard 2D-video technique successfully induces different emotional states, using more involving techniques, such as virtual reality goggles (see Mol, 2019) may help increasing the detectable effect of mood-congruency. Alternatively, it could be the case that the current affective state is not the most relevant cue for retrieval. In real life, self-relevant feedback can be embedded with negative emotions but with also other types of cues such that situations, images, sounds, etc. (Enke et al., 2019). Therefore, an alternative possibility to investigate the relative role of associative memory and self-enhancing memory in asymmetric recall of feedback would be to use such other types of context-dependent and not mood-dependent cues.

Overall, while we cannot rule out the existence of mood-congruency, it does not appear to be the dominant force. Instead, self-enhancement predictions are confirmed. These results underline the importance of motivational over affective factors in the formation of optimistic beliefs about the self. The relative dominance of self-enhancement offers direct implications for policies aimed at mitigating or removing biased judgments. Insofar as individuals mostly distort their memory because they consider negative feedback to be potentially harmful, removing *ex-ante* aversion to negative feedback should be the focus of this policy agenda.

Appendices

A1: Instructions *(translated from French)*

SESSIONS 1

(Instructions On paper)

Welcome!
Thank you for participating in this experience. By registering for this experience, you agree to participate in two sessions:

- a first session that is taking place today
- a second session to be held in one month, on [*date*], at [*daytime*].

Today, you will receive a fixed payoff of 5 euros for your participation in this first session. In one month, you will receive a fixed payoff of 15 euros for your participation in the second session. In addition to these 5+15=20 euros, you can earn extra money depending on the decisions you make during the experiment. The total of your additional earnings will be paid to you in cash in one month, together with the 15 euros.
In total, the experience is composed of 10 parts. In each part, you can earn extra money. At the end of the experiment, the computer will randomly select one of the 10 parts. Your payoff will correspond to the earnings made in the randomly selected part.
This experience is anonymous. The data generated in this experiment will only be used for scientific purposes. The researchers conducting this study are the only

people involved in the data collection and analysis process. This is true for both sessions of the experiment.
Throughout this session, it is forbidden to communicate with other participants and to use your mobile phones, tablets, etc.
The experiment will be conducted via computer. The experimenter will soon start the computer program. For each part of the experiment, you will be precisely instructed about your task and/or the decision problem you will be facing. For each part, you will receive specific instructions on your task and/or the decision problem you will face. Once you have read and understood the instructions, you can press the OK button to continue.
If you have any questions, you can press the red button on the left or right of your desktop at any time during the experiment. An experimenter will come to answer your question in private.

(Computerized Instructions)

We started by eliciting subjects' personal code. Subjects needed to enter the their day and month of birth plus the last four digits of their phone number. The resulting code was also elicited in the session one month later and allowed us to link participants' responses.

Please enter in the box below the day and month of birth of your father and mother, as well as your own day and month of birth. This code will be requested during the second session of the experiment in five weeks' time.
NEXT SCREEN

Before we begin Part 1, we would like to ask you a few questions. For each statement, please check the box that best describes your experience.
NEW SCREEN

(Emotional questionnaire)
For each statement, please check the box that best describes your feelings (0 means "not at all", 10 means "completely").
- Overall, are you feeling happy at the moment?
- Overall, are you satisfied with your life at the moment?
- Overall, are you feeling calm and/or relaxed at the moment?
- Overall, could you say that you feel at peace with yourself at the moment?
- Overall, could you say that you are feeling full of energy at the moment?
- Overall, are you feeling tired at the moment?
- Overall, are you feeling stressed and/or worried at the moment?
- Overall, could you say that you are often bored at the moment?
- Overall, are you feeling angry at the moment?
- Overall, do you often feel lonely?

NEW SCREEN

(Self-efficacy questionnaire)
For each statement, please check the box that best describes your experience.
- I can always manage to solve difficult problems if I try hard enough.
- If someone opposes me, I can find the means and ways to get what I want.
- It is easy for me to stick to my aims and accomplish my goals.
- I am confident that I could deal efficiently with unexpected events.
- Thanks to my resourcefulness, I know how to handle unforeseen situations.
- I can solve most problems if I invest the necessary effort.
- I can remain calm when facing difficulties because I can rely on my coping abilities.
- When I am confronted with a problem, I can usually find several solutions.
- If I am in trouble, I can usually think of a solution.
- I can usually handle whatever comes my way.

NEW SCREEN

Part 1
In this part, you are endowed with 10 euros. You can decide if you want to donate a part of these 10 euros to the French Red Cross. The French Red Cross helps people in different situations of need. It is especially active in areas of health and social emergency situations, but is also very engaged in providing help for refugees. You can decide if you want to donate some part of the 10 euros to the French Red Cross. You can donate every whole number amount up to 10 euros, or you can keep the 10 euros for yourself. If you decide to donate a positive amount, the experimenter will donate this amount to the French Red Cross after the experiment. The part of the 10 euros that you don't donate will increase your earning from this part of the experiment.

Which amount (in euros) would you like to donate?
NEW SCREEN

Part 2
In this part you receive 4 euros. Depending on your decisions you can earn additional money. In the following, you will go through a Raven IQ-test. This test is frequently used to measure intelligence. It is often found that performance in the test is associated with educational success and future income. The test consists of 10 tasks, and you have 10 minutes to solve it. You should try to correctly solve as many of the 10 tasks as possible.
NEW SCREEN

Subjects had to solve 10 Raven matrices.
NEW SCREEN

The exact same IQ test you just did was also conducted with a large number of participants who previously, exactly like you, participated in an experiment in the Gate-Lab. We randomly selected 9 of these participants. Together with these 9 participants you now form a group of 10 participants.
We constructed a ranking of this group based on performance in the IQ test. The group member that scored highest in the IQ test obtained rank 1. The group member with the second highest score obtained rank 2, etc... The group member with the worst performance in the IQ test obtained rank 10. In case of a draw between group members, the computer randomly decided who receives the higher rank.
In the following, we are interested in how you think you ranked in terms of IQ in the group of 10. We will ask you two questions. In both questions your earnings are higher, the more precise the estimate is you provide. The computer will randomly select one of the two questions, and this question will then be relevant for your earnings from this part of the experiment.
NEW SCREEN

First, we are interested in what you think is the likelihood (in percent) that you ranked in the upper half of the group. In other words, what do you think is the likelihood that in the group of 10, your rank is 5 or higher?
You will be paid based on the following formula: $Your\ payment\ (in\ euros) = 2 - 2(I(rank \leq 5) - beliefs = 100)^2$, where $I(rank \leq 5)$ is an indicator variable that takes the value 1 if your rank was in the upper half of the ranking, and p is your estimate in percent.
While this formula might look complicated, the basic idea is very simple. On average, your earnings are highest if you try to estimate as accurately as possible. In other words, the formula is such that it is best for you to provide an estimate that is as precise as possible. Your maximum earnings from your estimate are 2 euros, negative earnings are not possible.
On the next screen you can provide your answer.
NEW SCREEN

You can now enter your estimate. You can only enter whole numbers. The lowest possible number is 0 (percent). The highest possible number is 100 (percent).
I think the likelihood (in percent) that I rank in the upper half of the group of 10 is:
Subjects had to enter their estimate
NEW SCREEN

Your estimate of the likelihood (in percent) of ranking in the upper half of the group of 10 was: *number displayed*
Second, we are interested in how you would estimate the likelihood of holding specific ranks in the group of 10. We will first ask you to state an estimate for each of the 5 highest ranks. So what do you think is the likelihood that your rank is 1, what do you think is the likelihood that your rank is 2 etc., until rank 5.

Notice that the sum of the 5 estimates you provide must equal your estimate of ranking in the upper half of the groups of 10.
IMPORTANT: The sum of your 5 estimates must be equal to: *number displayed*
Afterwards we will ask you to state an estimate for each of the 5 lowest ranks. Notice that you will be paid for your estimates based on a similar formula as before. In case this question is payoff-relevant, the computer will randomly select one of the 10 ranks, and your earnings will be based on the following formula:
$Your\ payment\ (in\ euros) = 2 - 2(I(rank) - p = 100)^2$, where I(rank) is an indicator variable that takes the value 1 if this is indeed your rank, and p is your estimate for this rank in percent. So again, you should try to estimate as accurate as possible.
On the next screen you can enter your answers.
NEW SCREEN

You can now enter your estimates. You can again only enter whole numbers.
Again notice that the sum of your 5 estimates must be equal to: *number displayed*
Subjects had to enter their estimates
NEW SCREEN

Next, we are interested in how you would estimate the likelihood of holding each of the ranks 6-10. So what do you think is the likelihood that your rank is 6, what do you think is the likelihood that your rank is 7 etc., down to 10.
The sum of your 5 estimates must be equal to: *number displayed*
You can now enter your estimates below. You can again only enter whole numbers.
Subjects had to enter their estimates
NEW SCREEN

Part 3
In this part of the experiment, you receive 5 euros. Earlier you did a test to measure your intelligence. On the basis of your performance in the IQ test and the performance of 9 other randomly selected participants, we created a ranking. (You were, however, not informed about your position in this ranking.) We now randomly selected 3 out of the 9 other participants from your group. On the next screen we will inform you, for each of these 3 participants, whether you ranked higher or lower in terms of the IQ test.
NEW SCREEN

Of the 3 randomly selected participants from your group...
- Number of participants that ranked higher than you in terms of IQ: number displayed
- Number of participants that ranked lower than you in terms of IQ: number displayed

NEW SCREEN

Please repeat the feedback you just received. Of the 3 randomly selected participants from your group:
- How many ranked higher than you in terms of IQ? *subjects had to insert number*
- How many ranked lower than you in terms of IQ? *subjects had to insert number*

NEW SCREEN

We will now again ask you about the group consisting of yourself and the other 9 randomly selected participants. On the next screen, we will ask you how you now estimate your rank in this group in terms of IQ.

NEW SCREEN

What do you think now. What is the likelihood (in percent) that you ranked in the upper half of the group. In other words, what do you think is the likelihood that in the group of 10, your rank is 5 or higher?
You will again be paid based on the following formula:
$Your\ payment\ (in\ euros) = 2 - 2(I(rank \leq 5) - beliefs = 100)^2$, where I(rank 5 or higher) is an indicator variable that takes the value 1 if your rank was in the upper half of the ranking, and p is your estimate in percent. Again: While this formula might look complicated, the basic idea is very simple. On average, your earnings are highest if you try to estimate as accurately as possible. In other words, the formula is such that it is best for you to provide an estimate that is as precise as possible. Your maximum earnings from your estimate are 2 euros, negative earnings are not possible.
On the next screen you can provide your answer.

NEW SCREEN

You can now enter your estimate. You can only enter whole numbers. The lowest possible number is 0 (percent). The highest possible number is 100 (percent).
I think the likelihood (in percent) that I rank in the upper half of the group of 10 is: *Subjects had to enter their estimate*

NEW SCREEN

Part 4

In this part you receive 5 euros. Depending on your decisions you can earn additional money. Your task in this part is to count the number of zeros in tables. Once you have counted the number of zeros in a table, click OK. If you counted correctly, a new table will be generated. If you miscounted, you can try again twice. In other words, you have three tries per table. You receive 0.2 euros per correctly solved table. If you miscount a table three times, 0.2 euros will be deducted from your earnings. You have 4 minutes to count as many tables as you can.

Subjects had 4 minutes to work on the task.

NEW SCREEN

Part 5

In this part you receive 3 euros. Depending on your decisions you can earn additional money. You need to decide, how much money you want to invest in a lottery. You obtain an endowment of 200 cents. You can invest any amount between 0 and 200 cents in the lottery. The amount you choose not to invest will be directly added to your earnings. The lottery works as follows: The computer decides randomly if you win or loose in the lottery. The probability that you win is 1/3, the probability that you loose is 2/3. If you loose the lottery, you loose the amount you invested. If you win the lottery, the amount you invested will be multiplied by factor 2.5. This amount will then be added to your earnings from this part.
On the next screen, you can decide how much you want to invest.
NEW SCREEN

I would like to invest: *subjects could enter the investment amount.*
NEW SCREEN

Part 6
In this part of the experiment, we will ask you a series of question. In this part you earn 5 euros. On the next screen, the questions begin.
NEW SCREEN

We collected a number of socio-demographics, e.g., gender, age, field of study.
END

SESSIONS 2

(On paper)

Welcome!

On [*date of the experiment*], one month ago, you participated in an experiment. Today this experiment continues. One month ago you earned a fixed payment of 5 euros. For participating today, you obtain 15 euros. Thus, in total you receive a fixed payment of 20 euros. You can earn additional money. How much money you will earn depends on your decisions today and one month ago.
This experience is anonymous. The data generated in this experiment will only be used for scientific purposes. The researchers conducting this study are the only people involved in the data collection and analysis process. This is true for both sessions of the experiment.
Throughout this session, it is forbidden to communicate with other participants and to use your mobile phones, tablets, etc.
The experiment will be conducted via computer. The experimenter will soon start the computer program. For each part of the experiment, you will be precisely

instructed about your task and/or the decision problem you will be facing. For each part, you will receive specific instructions on your task and/or the decision problem you will face. Once you have read and understood the instructions, you can press the OK button to continue.
If you have any questions, you can press the red button on the left or right of your desktop at any time during the experiment. An experimenter will come to answer your question in private.

(Computerized)

We started by again eliciting subjects' personal code to be able to match responses between the two dates. Subjects needed to enter the day and month of their birth-date plus the last four digits of their phone number.

Please enter in the box below the day and month of birth of your father and mother, as well as your own day and month of birth. This code will be requested during the second session of the experiment in five weeks' time.
NEXT SCREEN

Before we starting the different parts of the experiment, we would like to ask you a few questions. For each statement, please check the box that best describes your experience.
NEW SCREEN

(Emotional questionnaire)
NEW SCREEN

(Self-efficacy questionnaire)
NEW SCREEN

Now, we would like you to watch a video clip. The video clip will appear on your screen. Please make yourselves comfortable: the clip will last about 4 minutes.

Please click "next" to start watching the video clip.
NEW SCREEN

Now please indicate on the two manikin-like scales below how you emotionally reacted to the video clip.
The overall emotional experience I felt while watching the video clip was ...
NEW SCREEN

The overall emotional experience I felt while watching the video clip was ...
NEW SCREEN

Please indicate the emotion that best captures what you have felt while watching the movie clip. Only one option allowed: The movie clip made me feel...
- Afraid/Scared/Anxious
- Bored/Jaded/Uninterested
- Neutral (no emotional reaction)
- Excited/Eager/Enthusiastic
- Sad/Gloomy/Depressed
- Calm/Relaxed/Peaceful

NEW SCREEN

Part 7

In this part you receive 2 euros. Depending on your decisions you can earn additional money. Part 7 consists of 2 subparts. In case part 7 will be payoff-relevant, one of the two subparts will be randomly selected and will determine your earnings.

NEW SCREEN

We would like to know which parts of the experiment one month ago you remember. Please try to summarize each part you remember in one sentence. For each sufficiently accurate description of a part (as evaluated by the experimenter), you earn 0.5 euros.

NEW SCREEN

In the following, please try to describe the different parts of the experiment 1 month ago (1 sentence for each part). In front of you, you find a sheet of paper. Please write down your descriptions on the sheet of paper. Please inform the experimenter once you are finished.

The experiment continued once all subjects had handed in their answer sheets.

NEW SCREEN

As a reminder: 1 month ago you participated in an IQ test. We had conducted the exact same IQ test you did with a large number of participants who previously, exactly like you, had participated in an experiment in the Gate-Lab. We had randomly selected 9 of these participants. Together with these 9 participants you formed a group of 10 participants.

We had constructed a ranking of this group based on performance in the IQ test. The group member that scored highest in the IQ test obtained rank 1. The group member with the second highest score obtained rank 2, etc... The group member with the worst performance in the IQ test obtained rank 10. In case of a draw between group members, the computer randomly decided who received the higher rank.

We had randomly selected 3 out of these 9 participants, and informed you, for each of the 3 participants, whether you ranked higher or lower in terms of IQ.

In the following, we would like to know if you remember how you ranked compared to the three randomly selected participants. If you answer correctly, you receive 2 euros. On the next screen you can provide your answer.
NEW SCREEN

Of the 3 randomly selected participants from your group: How many ranked higher than you in terms of IQ?
NEW SCREEN

Part 8
As a reminder: 1 month ago you were endowed with 10 euros and were given the possibility to give part of this amount to the French Red Cross.
In the following, we would like to know if you remember how you ranked compared to the three randomly selected participants. If you answer correctly, you receive 2 euros. On the next screen you can provide your answer.
NEW SCREEN

Out of 10 euros, how much did you give to the French Red Cross?
NEW SCREEN

Before the end of the experiment, we would like to ask you some questions. On the next screen, the questions begin.
NEW SCREEN

Subjects were asked to guess the purpose of the study, and to report on a 10-item scale how much confident they were with each recall task (IQ feedback and donation) and how much effort they provided to retrieve the information at stake. Then, the experiment ended.
END

A2: Tables

Table A2.1: Summary statistics - Participants, by treatment

	Treatment		
	Neutral	Positive	Negative
Male	62%	50%	51%
Age	24	26	25
Students	17%	13%	13%
Nb. neg. comparisons	1.71	1.45	1.58
Rank in IQ test	5.84	5.42	5.61
Nb. correct raven	6.07	6.28	6.25
Ave. nb. of subjects per session.	16	15	17.25

Note: this Table reports the results of two-tailed M-W tests and two-sample tests of proportions in which each individual is taken as an individual observation. Positive and Negative treatments are compared to the Neutral treatment

Table A2.2: Reported emotional state and self-esteem *before* treatment manipulation

	Affective state			Self-esteem		
	Neu.	Pos.	Neg.	Neu.	Pos.	Neg.
Session 1	18.65	17.33	14.86	34.23	35.50	34.38
	(16.65)	(18.63)	(19.36)	(4.16)	(5.35)	(4.39)
Session 2	16.68	16.47	15.01	34.48	35.27	34.23
	(16.26)	(18.16)	(19.36)	(4.86)	(5.60)	(5.25)
Diff.	1.98	0.87	-0.12	-0.25	0.23	0.14
	(14.41)	(15.68)	(14.98)	(3.09)	(3.33)	(3.79)

Note: The difference between Session 1 and Session 2 is never significantly different from 0 (Wilcoxon signed-rank tests). None of the reported value are significantly different from each other, except the average reported self-esteem in session 1 between the Neutral and Positive treatment (p=0.073, Mann Whitney test). Standard deviations are in parentheses. The affective state can range from -50 to 50. The self-esteem can range from 11 to 44.

Chapter 3

Motivated Memory of Unethical Decisions[1]

> I think as you go through life you do things that you forget about them, and you try to look back and say: "Well, I was a saint, I was good, everything I did was good, everything I thought was good."
>
> Jack Abramoff, *In It To Win: The Jack Abramoff Story*

3.1 Introduction

In 2007, Alberto Gonzales, then US Attorney General under George W. Bush's presidency, was involved in the arbitrary firing of several U.S. Federal Attorneys. During the Senate hearing, set up to investigate his role in the case, he claimed amnesia and uttered the sentences "I don't recall", "I have no recollection" or "I have no memory" more than 60 times. In 2012, rapper Lily Wayne declared to the Court not recalling any of the criminal actions that were pending against him.

[1]This chapter is a joint work with Fabio Galeotti and Marie Claire Villeval.

Examples of crime-related amnesia like this abound in legal judgments. Claims of amnesia have been found to occur in murders (Taylor and Kopelman, 1984), sexual harassment (Bourget and Bradford, 1995), domestic violence (Swihart et al., 1999) or fraud (Kopelman et al., 1994). Often, crime-related amnesia is used as an attempt to avoid responsibility for past misdeeds or to impede police investigation. However, it can also be genuine amnesia in the sense that an offender is unable to correctly retrieve memories of his or her past offenses (Cima et al., 2002).

Memory impairment is not only a prerogative of criminals and it does not only concern serious offenses. Indeed, recent studies in psychology show that people tend to forget the details of their past unethical behavior, so that they can think of themselves as honest persons (Kouchaki and Gino, 2016; Stanley et al., 2017). People care about being moral and exhibit perceived cheating aversion (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Khalmetski and Sliwka, 2019; Abeler et al., 2019); but sometimes, they engage in actions that may contradict their desire for moral image. Forgetting may be used as a strategy to restore consistency between their past actions and their demand for positive self-image. This idea is well captured by the opening quote made by former lobbyist and convicted felon Jack Abramoff during a 2012 interview at the University of Texas at Austin while talking about his misdeeds.

What can individuals do to preserve their self-image after acting immorally? Economists and psychologists have highlighted different strategies that people adopt to preserve their self-image. For instance, individuals can avoid knowing the consequences of their behavior (Feiler, 2014; Grossman and Van Der Weele, 2017), exploit norm-uncertainty about lying behavior to justify their own decision to lie (Bicchieri et al., 2019), claim that they have changed (Stanley et al.,

2017), shift the blame onto someone else (Bartling and Fischbacher, 2011; Oexl and Grossman, 2013), balance their moral behavior over time (Nisan and Kurtines, 2013; Ploner and Regner, 2013; Gneezy et al., 2014; Cojoc and Stoian, 2014), or use narratives consisting in, for example, downplay the externalities and/or reinterpret the circumstances of their actions (Bénabou et al., 2018). Another strategy that individuals may use to preserve their moral self-image is to forget or manipulate, either consciously or unconsciously, unwanted memories. This process of forgetting or distorting the memory is called motivated memory (Singer and Salovey, 1996). Motivated memory allows people to directly derive utility from thinking of themselves in good terms. Beyond this purely hedonic motivation, motivated memory may also be used as an instrument to justify future possibly unethical actions. For example, if individuals are able to remember the eco-friendly actions they sometimes undertake but systematically forget the environmentally irresponsible ones, not only they will have a clean conscience from thinking of themselves as eco-friendly persons (and derive a positive utility from it), but they can also use these positive memories to justify future irresponsible acts.

In this study, we investigate the existence of motivated memory in the context of dishonest decision-making. Despite the vast literature on unethical behavior that has flourished in the last decade (for surveys, see Rosenbaum et al. 2014; Irlenbusch and Villeval 2015; Jacobsen et al. 2018; Abeler et al. 2019), there is no economic study investigating memory manipulation as a self-management mechanism to sustain moral self-image when acting dishonestly. We explore not only whether individuals forget their past unethical behavior to sustain their desire for moral self-image, but also whether they manipulate their memory as an excuse not to engage in subsequent morally responsible behavior (in our experiment, giving

back undeserved money).[2] Indeed, motivated memory can be thought of as an instrument, i.e., as something that is not only valued for *its own sake* (the hedonic value of keeping a good image after a misconduct), but that also *helps to get something else* that is valued for its own sake (keeping undeserved money). In our study, memory manipulation can be used both as "post-violation justifications [that] alleviate the *experienced* threat to the moral self" (Shalvi et al. 2015, p.1), when participants forget that they have cheated in the past, **and** as "pre-violation justifications [that] lessen the *anticipated* threat to the moral self" (Shalvi et al. 2015, p. 1), when dishonest participants plan to keep undeserved money which they are expected to give back. In this respect, we provide the first experimental test of the impact of anticipated decisions on memory manipulation.

To study whether individuals (i) manipulate the memory of past dishonest choices, and (ii) use their memory as an instrument to justify their future decisions, we conducted an on-line experiment that was divided into two parts separated by three weeks. In the first part, participants were asked to play a repeated "wheel game" — a modified version of the "mind game" (see Jiang, 2013; Shalvi and De Dreu, 2014; Potters and Stoop, 2016; Gneezy et al., 2018) — in which participants could misreport their outcomes to increase their payoff at no risk of detection and sanction. The experimenter could infer whether a participant was dishonest or not by comparing the distribution of the individual reports with a uniform distribution but was not able to prove that a report was a lie. Three weeks later, participants were asked to recall the distribution of the outcomes that they reported in the first part, and were paid for the accuracy of their recalls. This design allowed us to measure participants' memory errors, defined as the differ-

[2]Our study strongly differs from the few ones in psychology that only investigate memory biases as a *consequence* of past unethical behavior (Shu et al., 2011; Kouchaki and Gino, 2016; Stanley et al., 2017). These few studies in psychology rely on attitudinal measures of memory rather than focusing on behavior, and do not incentivize participants for providing truthful recalls.

ence between the average *reported* outcome in the first part of the experiment and the average *recalled* outcome in the second part, and to correlate the magnitude of the memory error with the dishonesty of the participants. Systematic positive memory errors capture motivated memory. Moreover, we varied between-subjects whether participants were given or not the possibility to reduce their payoff by a fixed amount at the end of the second part of the experiment. In the Hedonic treatment, participants were not given this option. Hence, the only reason why a dishonest participant could exhibit motivated memory was to maintain a moral self-image. In the Instrumental treatment, participants were given the possibility to reduce their payoff and they were especially encouraged to do so if they had misreported many numbers to their advantage in the wheel game they performed three weeks before. Therefore, it was made salient that dishonest participants were expected to reduce their payoff. In this treatment, dishonest participants could thus motivate their memory not only for hedonic reasons (i.e., preserving a moral image) but also for instrumental reasons (i.e., not giving back part of the undeserved money). In both treatments, honest individuals had no reason to manipulate their memory since accurate recalls were incentivized. So, any memory errors on their side could be attributed to random rather than motivated forgetting, and these errors should be distributed around zero.

It is probably easier to recall a uniform distribution rather than a biased distribution. Thus, in the Hedonic and Instrumental treatments, dishonest individuals could display positive memory errors not because of motivated memory but because they had to recall a different distribution than the one faced by honest individuals. To control for this, we ran Hedonic-Control and Instrumental-Control treatments. These treatments are similar to the Hedonic and Instrumental treatments, respectively, except that participants were not allowed to cheat in the first

part of the experiment: they could not enter a number different from the actual number displayed in the wheel. Moreover, the numbers they had to report reflected the aggregate distribution of the numbers reported in the Hedonic and Instrumental treatments. In other words, these participants could observe uniform or non uniform distributions. In these Control treatments, participants had no reason to manipulate their memory. Yet, memory errors could be positive if forgetful participants had a tendency to recall a uniform distribution. The comparison of the Hedonic and Instrumental treatments with the Control treatments allowed us to separate motivated memory from non-motivated forgetting. The comparison of the memory errors in the Hedonic-Control versus Instrumental-Control treatment additionally allowed us to control for the mere reaction to the possibility to give money back, when no moral image is at stake. Since participants have no reason to manipulate their memory (in both treatments they were not allowed to cheat in part 1), we expect no difference in memory errors between the Hedonic-Control treatment and the Instrumental-Control treatment.

Our results show that hedonic considerations — i.e., recalling oneself as an honest person — were not sufficient to trigger a significant memory manipulation in our setting. Dishonest individuals in the Hedonic treatment did not recall their past decisions less accurately than participants in the Control treatments. By contrast, when memory manipulation had an instrumental value — i.e., when forgetting past lies could serve as a justification *not* to give undeserved money back — dishonest individuals did recall their past behavior with less accuracy than participants in the Hedonic and Control treatments, although this reduced their payoff in the recalling task. This finding suggests that individuals recalled selectively as a self-excusing strategy to justify anticipated future decisions. This confirms that memory is involved in the various strategies people use to motivate their beliefs

about themselves and justify they can behave immorally while keeping a positive self-view.

The remainder of the paper is organized as follows. Section 2 briefly reviews the related literature. Section 3 presents the experimental design and procedures. Section 4 outlines the behavioral conjectures. Section 5 reports the results and section 6 concludes.

3.2 Related Literature

Our study contributes to two strands of the literature. First, it contributes to better understand moral reasoning when facing an opportunity to misbehave. Second, it contributes to the recent but growing economic literature on motivated memory. This section briefly reviews these two strands of the literature and illustrates how our study contributes to each of them.

Many people sometimes act unethically without considering themselves as dishonest or immoral persons (Shalvi et al., 2015). The recent economic literature on cheating and lying has tried to identify when and to which extent people lie, and it has explored various strategies used by individuals to preserve their moral-self while acting dishonestly.

Regarding the existence and extent of dishonest behavior, a large body of experimental evidence shows that even in the absence of risk of detection and sanction, not all people lie and when they do, most do not lie to the full extent (Abeler et al., 2019). This was found both in the lab when participants misreport the outcome

of a private random draw or overstate their performance in a task (e.g., Mazar et al. 2008; Shalvi et al. 2011; Abeler et al. 2019; Gneezy et al. 2018; Kajackaite 2018), and in the field when individuals fraud in public transportation (Dai et al., 2017), keep undeserved money (Azar et al., 2013; Potters and Stoop, 2016), do not return misdirected letters containing cash (Keizer et al., 2008; Franzen and Pointner, 2013; Andreoni et al., 2017) or keep found wallets containing money (e.g., Cohn et al., 2019).[3] In a meta analysis combining data from more than 90 experimental studies, Abeler et al. (2019) found that individuals forgo on average about 75% of the potential gains by not lying maximally.

When individuals misbehave, they can experience intrinsic costs from behaving dishonestly, from being potentially perceived as dishonest persons, and from the violation of social norms (Dufwenberg and Dufwenberg, 2018; Gneezy et al., 2018; Abeler et al., 2019; Khalmetski and Sliwka, 2019). While these costs can prevent individuals from acting dishonestly and reveal a preference for truth-telling, they may also trigger strategies used by individuals to behave self-interestedly while keeping a good image. Identifying these strategies has been the focus of several experimental studies in economics and psychology. Using dictator games, Dana et al. (2007) found that participants exploit the uncertainty of the positive outcomes in order to behave more selfishly. Feiler (2014) and Grossman and Van Der Weele (2017) showed that dictators tend to avoid knowing the consequences of their action on the receivers, while Oexl and Grossman (2013) reported that dictators tend to shift the blame of unfair outcomes onto an intermediary. Closely related to dishonesty, Stanley et al. (2017) showed that individuals judge their own actions as more morally wrong for periods of the past when they considered they were different persons than for periods of the past when they considered that

[3] see Rosenbaum et al. (2014); Irlenbusch and Villeval (2015); Jacobsen et al. (2018); Abeler et al. (2019) for surveys on unethical behavior.

they were very similar to whom they are today. Differently from these studies, we explore another strategy that individuals may use to preserve their moral self-image: motivated memory. We investigate whether individuals manipulate their memory to lower the moral cost associated to dishonesty. In particular, we test whether individuals use forgetting of past dishonest decisions to disguise their lies to themselves and, thereby, maintain a positive self-image.

Few theoretical papers in economics tried to model motivated memory.[4] Bénabou and Tirole (2002) modelled memory manipulation as the outcome of a game played between the multiple selves of an individual. In equilibrium, the individual forgets information that undermines his long-term goals (forgetting for instrumental reasons) or lowers his self-esteem (forgetting for hedonic reasons). Building on Bénabou and Tirole (2002)'s self-deception framework, Gottlieb (2014) shows that after observing a negative signal, the decision-maker faces a conflict between forgetting the signal in order to have a better self-image but making a less appropriate decision, or recalling it in order to make a better decision. When there is no ex-post decision to make (i.e., when forgetting has only hedonic value), the self-image factor takes over and the decision-maker recalls a negative signal with probability below the actual percentage. When there is an ex-post decision and the self-image and decision-making factors have opposite signs, the amount of memory manipulation depends on the difference between the marginal benefit and the marginal cost of remembering.

Few experimental studies in economics have investigated the use of memory as a self-deceptive mechanism to sustain positive self-image. Moreover, while the-

[4]The economic literature modelling cognitive limitation in recalls and their impact on belief formation and decision-making is substantial (Dow, 1991; Piccione and Rubinstein, 1997; Mullainathan, 2002; Bénabou and Tirole, 2004; Brunnermeier and Parker, 2005; Bénabou and Tirole, 2006; Wilson, 2014; Bordalo et al., 2017). However, they all consider memory as an imperfect or limited faculty of people but not as a self-deceptive strategy used by individuals to sustain their demand for positive self-image.

oretical models recognize both the hedonic and instrumental values of motivated memory, the experimental literature has focused only on the hedonic motives of motivated memory. Chew et al. (2018) showed that individuals exhibit asymmetric recalls of past performance in an IQ test. Individuals forget more their incorrect answers than their correct ones (selective amnesia) but they also exhibit false memory, encompassing delusion (remembering a correct answer when there was none) and confabulation (transforming a wrong answer into a correct one). Zimmermann (2019) also provided evidence of asymmetry in the recall of feedback on past relative performance in an IQ test. To study the dynamics of motivated beliefs, he considered different treatments where he manipulated the incentives for correct recalling and the time between feedback and the elicitation of beliefs. Zimmermann (2019) showed that individuals manage to suppress negative feedback that threaten their desire to view themselves as intelligent persons. Individuals perfectly encoded feedback but recalled them asymmetrically one month later. Using a word-entry task instead of an IQ test, Li (2017) showed that having to forecast performance and receiving feedback both eliminate biased recalls. In the social domain, Li (2013) investigated recollection of decisions in a trust game by manipulating the delay between the decisions and the recollection. While he found asymmetric recalls among trustors (betrayed trustors recall less accurately their decisions than trustors who benefited from reciprocity), there was no clear evidence of selective memory among trustees. Saucet and Villeval (2019) showed that dictators remember more their altruistic than their selfish choices. A causal effect of the responsibility of the decisions was identified, as the recall asymmetry disappeared when options were selected randomly by the computer program. Carlson et al. (2018) showed that people misremember the extent of their selfishness when their actions fall short of their own personal standards. While these studies investigate motivated memory driven by self-image concerns in the domain of intel-

lectual ability or social preferences, we focus on the memory of dishonest decisions.

More closely related to our study are two recent papers in psychology, by Stanley et al. (2017) and Kouchaki and Gino (2016). In Stanley et al. (2017), participants were asked to recall specific events from their personal pasts which involved lying. Stanley et al. (2017) found that participants rated as less morally wrong remembered events in which they were actor of the action than remembered events in which they were the recipient of the action. Kouchaki and Gino (2016) showed that memories of unethical actions gradually became less clear and vivid than memories of ethical actions. They called this phenomenon "unethical amnesia": people forget the details of their unethical actions, so that they can keep thinking of themselves as honest individuals. Kouchaki and Gino (2016) also found that people who cheated a first time and forgot their past behavior were more likely to cheat a second time. While we also investigate individuals' recall of honest *vs.* dishonest actions, in contrast to these studies we do not only focus on motivated memory as a way to restore moral self-image but also as an excuse to justify future behavior.[5] In our study, participants know, *before* recalling, that they may be tempted to keep undeserved money. Therefore, they can use unethical amnesia as an excuse not to bend their moral self in the future.[6]

[5]In Kouchaki and Gino (2016), participants were *not* informed of the second cheating task when being asked to recall their behavior in the first cheating task. Therefore, they could not, at any time in the experiment, use forgetting as an excuse to cheat a second time.

[6]We also differ from Stanley et al. (2017) and Kouchaki and Gino (2016) in that i) we elicit memory of the decision itself and not of the environment surrounding it; and ii) we incentivize correct recalls, which creates a trade-off between manipulating one's memory and losing money from inaccurate recall, or recalling accurately and incurring potential moral cost from threatened self-image and unjustified future questionable decisions. Stanley et al. (2017) control for difference in memory by asking participants to rate on a 7 item-scale how well they remembered each event. Kouchaki and Gino (2016) measure memory using items adapted from the Memory Characteristics Questionnaire (MCQ) and the Autobiographical Memory Questionnaire (AMQ), which assess various qualitative characteristics of one's memory, notably clarity (how vividly a person remembers an event) and thoughts and feelings (how a person remembers the feelings and thoughts experienced during the event), also on item-scales. In none of the cases correct recalls were incentivized.

3.3 Experimental Design and Procedures

We describe the design of the experiment and then detail the procedures.

3.3.1 Experimental Design

Our experiment is divided into two parts separated by three weeks. In the first past, participants played an original wheel game allowing them to misreport the outcome of a random draw. Then, participants performed a word recognition task to control for participants' general memory capacity. In the second part that took place three weeks later, participants were asked to recall the distribution of their reports in the past wheel game. Instructions are included in Appendix 1. We now describe each part in detail.

Part 1

Wheel Game

In the first part of the experiment, participants played an original "wheel game". A wheel with six empty squares was displayed on the participants' screens (see Figure 3.1). Participants had to choose one square. Once they had made a choice, the program randomly displayed a number between 1 and 6 (included) in each square of the wheel. Each number appeared only once in the wheel. Participants were asked to report the number displayed in the square that they had previously chosen.

There were two methods to choose the square in the wheel. Participants were assigned randomly either the "mind" or "click" method. If participants were assigned the "mind" condition, they had to choose one square of the wheel *in their head* before the numbers were displayed on the wheel. This means that the square

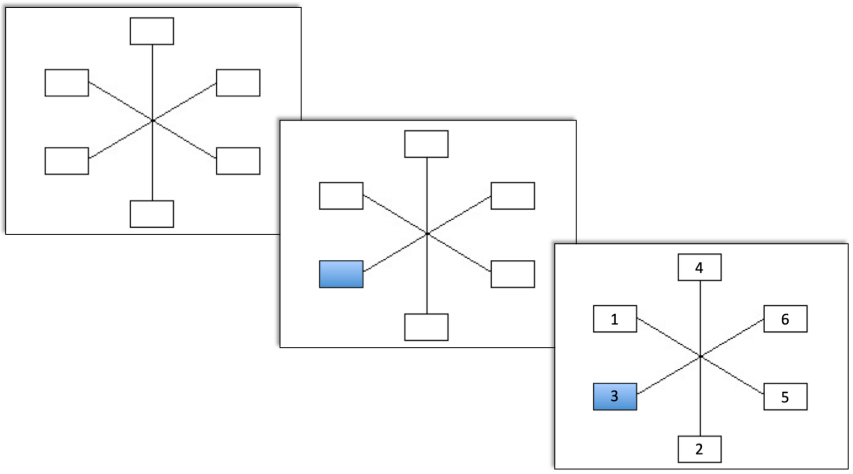
they chose was not observable by the experimenter and they could misreport the number appearing in the chosen square. In this condition, the wheel game is a mind game (see e.g., Jiang, 2013; Shalvi and De Dreu, 2014; Potters and Stoop, 2016; Gneezy et al., 2018), with multiple outcomes. If participants were assigned the "click" condition, they had to choose one square of the wheel *by clicking on it*. The selected square was then highlighted in blue. This means that the square they had chosen was observable by the experimenter. Participants in this condition could not misreport their outcome. If they reported a different number than the number displayed in the square they had chosen, they received a pop-up message saying that the entered number was not the right one.[7] Participants were not informed that they would be asked to recall the distribution of their reports in the second part of the experiment three weeks later.

Participants played the wheel game 20 times. At the end of the second part of the experiment, one wheel was selected at random among the 20. Participants' earnings depended on the number that they reported for that wheel. They earned \$0.10 if they had reported "1", \$0.20 if they had reported "2", \$0.30 if they had reported "3", and so on and so forth, up to \$0.60.

Recognition Task

Once they had played the wheel game, participants had to perform a word recognition task based on Roediger and McDermott (1995). This task was used to control for participants' general memory capacity. Participants were presented with five lists of words. Each list contained six words belonging to the same lexical field. Each word was displayed on the screen one by one for less than one second. Partic-

[7]The wheel game is close to the game used by Gneezy et al. (2018) in their Numbers treatment in which participants were asked to click, in private, on one of ten boxes on a computer and reveal an outcome. In our game participants choose one of six squares on a computer, either in their mind or by clicking on it according to treatments.



Notes: The Figure displays the wheel game played by participants in part 1. A wheel with six empty squares was displayed on the screen (top-left wheel). Participants chose one square of the wheel (middle wheel, square highlighted in blue for illustration). When they had made their choice, the program then displayed a number between 1 and 6 (included) in each square of the wheel (bottom-right wheel). Participants were asked to report the number displayed in the square that they had chosen.

Figure 3.1: Wheel game

ipants were informed, before observing these words, that at the end of the session they would have to recognize, among 35 words, some of the words that had been shown to them earlier. More specifically, for each of the 35 words, they would have to indicate whether it had been presented before. If they remembered having seen the word before, they had to check the button "Old". If they did not remember having seen the word, they had to check the button "New". The order of the 35 words was the same for all participants.

Participants earned \$0.02 for each correct answer. After observing the five lists of words, but before the recognition task, participants were asked to complete a demographic questionnaire.

Part 2

Part 2 took place three weeks later. At the beginning of this part, participants were informed that they might be given or not the possibility to reduce their payoff by \$0.75 at the end of the experiment, and whether they had or not this opportunity was determined randomly. If participants were given this possibility, they would have to decide at the end of the experiment whether they were willing or not to reduce their payoff by \$0.75. We refer to this case as the "Instrumental" condition.[8] If participants were not given the possibility to reduce their payoff by \$0.75, they would have no decision to make at the end of the experiment and their payoff would remain intact. We refer to this case as the "Hedonic" condition. Participants were also informed that those who would be given the possibility to reduce their payoff by \$0.75 were especially encouraged to do so if they misreported many numbers to their advantage in the wheel task they performed three weeks earlier. Precisely, they received the following instructions: *"At the end of today's part, you may be given or not the option to reduce your total earnings by \$0.75. It would be nice to reduce your total earnings if you are given this option and you have misreported numbers to your advantage several times in the wheel task that you performed three weeks ago"*. Therefore, participants that had been dishonest in the first part of the experiment and that were given the possibility to reduce their payoff had the opportunity to give back money that they had earned unethically. Importantly, *all* participants received the same set of instructions and were told about the two conditions in advance. Informing *only* participants from the Instrumental condition that dishonest players were expected to reduce their payoff would have made salient that cheating was somehow immoral *only* to these participants. By contrast, informing all participants of it avoided creating any

[8]This denomination was not used in the instructions.

difference in moral cost from remembering past unethical behavior between the Hedonic and Instrumental conditions.

Recall Task

After being informed whether they actually had or not the possibility to reduce their payoff, but *before* deciding whether to do it or not, participants performed the recall task. After reporting whether they remembered or not the wheel task they performed in part 1, they were asked to recall the distribution of the 20 numbers they reported in the wheel task. Precisely, they had to recall how many times they reported the numbers "1", "2", "3", "4", "5", and "6". Participants were paid depending on the accuracy of their recalls. At the end of the experiment, one number between 1 and 6, inclusive, was selected at random. Participants received \$1 if their recall was accurate, i.e., if they recalled exactly how many times they reported that number three weeks earlier. Participants received \$0.5 if their recall was inaccurate by plus or minus one. Otherwise, they earned zero.

Once they had completed the recall task, participants allocated to the Instrumental condition had to decide whether they were willing to reduce or not their total earnings by \$0.75. Finally, all participants completed questionnaires related to guilt and religiosity.

3.3.2 Treatments

Our 2x2 between-subject design is summarized in Table 3.1. We varied whether or not participants could cheat in part 1 and whether or not they could reduce their payoff in part 2. The condition in which participants could not cheat in the wheel task (the click condition) and were not offered to reduce their payoff in part 2 is called the Hedonic-Control treatment. In this treatment, participants have

no reason to exhibit unethical amnesia, either for hedonic motives (since moral self-image is not at stake when recalling), or for instrumental motives (they do not need excuses to not return money). This baseline treatment controls for imperfect memory, i.e., for memory errors when there is no intrinsic incentives for memory manipulation.

The Instrumental-Control treatment is similar to the Hedonic-Control treatment, except that participants are given the possibility to reduce their payoff by \$0.75 at the end of the experiment. The comparison between the Instrumental-Control and Hedonic-Control treatments controls for the mere reaction to the possibility to give money back. Since participants have no reason to manipulate their memory (since in both treatments they were not allowed to cheat in part 1), we expect no difference in memory errors between the Hedonic-Control treatment and the Instrumental-Control treatment.

The Hedonic treatment is similar to the Hedonic-Control treatment, except that participants could misreport their outcomes in part 1. The comparison between the Hedonic and Control treatments indicates whether individuals manipulate their memory in part 2 for hedonic reasons. If in the Hedonic treatment it is morally costly for dishonest players to recall that they have cheated in the first part of the experiment, we should observe higher memory errors in the Hedonic treatment than in the Control treatments.

The Instrumental treatment is similar to the Hedonic treatment, except that participants were informed, *before* recalling, that they would be given the possibility to reduce their payoff by \$0.75 later in the experiment. Importantly, they were encouraged to do so if they have misreported many numbers to their advantage in the wheel task. The comparison between the Hedonic and Instrumental treatments indicates whether or not participants exhibit more unethical amnesia when it does not only serve to make oneself look better (hedonic motive) but also to justify not

giving back part of the money their had earned unethically (instrumental motive). If in the Instrumental treatment dishonest players use forgetting of past cheating behavior as an excuse *not* to reduce their payoff, we should observe higher memory errors in the Instrumental treatment than in the Hedonic treatment.

To be able to compare the recalls of participants across treatments, it is important that participants in the Control treatments (Hedonic-Control and Instrumental-Control treatments) have the same aggregate distribution of numbers to recall than participants from the Hedonic and Instrumental treatments. Thus, we first ran the Hedonic and Instrumental treatments and we replicated the aggregated distribution of reported numbers from these treatments in the Control treatments. We calculated the relative frequency of each reported number in the main treatments and we assigned the same probability of occurrence of each number in the Control treatments.

Table 3.1: Summary of the treatments

		Part 2	
	Participants	**Cannot** reduce	**Can** reduce
Part 1	**Can** cheat	*HEDO*	*INSTRU*
	Cannot cheat	*HEDO-Control*	*INSTRU-Control*

Notes: In HEDO, participants can cheat in part 1 but are not given the possibility to reduce their payoff in part 2. In INSTRU, participants can cheat in part 1 and are given the possibility to reduce their payoff in part 2. HEDO-Control and INSTRU-Control replicate HEDO and INSTRU, respectively, except that participants are not allowed to cheat in part 1.

3.3.3 Procedures

The experiment was programmed using Java language and conducted on Amazon Mturk. A total of 1550 participants in the U.S. were recruited for the first part

of the experiment. Three weeks later, the same participants were invited to complete the second part. Overall, 1322 participants completed the second session (85.29%).[9] 488 participated in the Hedonic treatment, 508 in the Instrumental treatment, 163 in the Hedonic-Control treatment and 163 in the Instrumental-Control treatment.[10] Table A2.1 in Appendix 2 summarizes the participants' characteristics in each treatment.

Instructions for each part were self-contained and displayed on the participants' screen as the experiment processed (see Appendix 1). No feedback on performance or earnings was provided until all parts were completed. The first part lasted on average 9.5 minutes. The second part lasted on average 6 minutes. At the end of the first part, participants only received $1.5 fixed payoff for their participation. At the end of the second part, participants received $1.5 fixed payoff for completing the second part, plus the joint earnings made in the first and in the second parts. Participants were informed of this payment scheme at the very beginning of the experiment. Participants earned on average 4.31US$ (S.D. 0.48) and were paid on their Mturk account within 48 hours after the each part of the experiment.

3.4 Behavioral Conjectures

The following section formulates two behavioral conjectures regarding participants' recalls conditional on (i) whether they could cheat or not in part 1, and (ii) whether they were offered or not the possibility to reduce their payoff in part 2.

[9] The attrition rate (14.71%) was consistent with the average attrition rate of 15% observed by Amazon Mturk for longitudinal studies.

[10] The sample size was determined by following standard practices in the literature, see Appendix 6 for details about the sample size calculation.

When performing the wheel task in part 1, participants in the main treatments face a trade-off between reporting honestly each number but making less money than if they lied, or over-reporting numbers to increase their payoff but incurring potential moral costs from lying or being perceived as a liar. At the time of the decision, the temptation of making more money may take over and part of the participants may misreport numbers even though it comes with bending their own morality. At the time of the recollection, however, they may recall that they were honest or, at least, not that dishonest.[11] When participants have been honest, recalling correctly the reported numbers has no undesirable implication in terms of self-image. However, when participants have been dishonest, recalling correctly past cheating behavior may be self-image threatening. Therefore, to lower the moral cost from recalling that they have been dishonest, participants might minimize the extent of their lies by recalling a lower frequency of reported high numbers in session 1.[12] By contrast, in the Hedonic-Control treatment participants could not choose the numbers to report and were thus not responsible for their value. For these participants, recalling correctly the reported numbers has thus no moral implication for the self, and this, regardless of whether they have to recall a honest or a dishonest distribution. Therefore, we expect that participants identified as dishonest players in the Hedonic treatment will exhibit higher positive memory errors than participants who had to recall a dishonest distribution in the Control treatments. By contrast, we do not expect any difference in memory errors between honest participants in the Hedonic treatment and participants who had to

[11] Such temporal inconsistencies have been modeled by Tenbrunsel et al. (2010) in a "want/should" framework to explain bounded ethicality. During the action phase, the "want" self −characterized by a relative disregard for ethical considerations− dominates. During the prediction and recollection phases, however, the "should" self −characterized by intentions and beliefs on how one ought to behave− dominates.

[12] This is consistent with Bénabou and Tirole (2002) where individuals are motivated to forget signals that lower their self-esteem and with Gottlieb (2014) where the decision-maker, when he has no ex-post decision to make, recalls negative signals that threat his self-image with probability below the natural percentage (imperfect memory).

recall a honest distribution in the Control treatments. We state our first conjecture as follows:

Conjecture 1 (Hedonic value of memory) *Dishonest individuals in the Hedonic treatment exhibit higher positive memory errors than participants that had to recall a dishonest distribution in Control treatments. No difference is expected between honest individuals and participants that had to recall an honest distribution.*

In the Instrumental treatment participants were informed, *before recalling*, that they would have the possibility to reduce their payoff by \$0.75 later in the experiment. They were especially encouraged to do so if they had misreported many numbers to their advantage in the wheel task. Precisely because of this information, honest and dishonest individuals may not experience the decision to reduce their payoff the same way. On the one hand, an individual that reported truthfully the numbers in the wheel task should be perfectly fine with keeping the money since he does not feel concerned (he has no sins to redeem from). On the other hand, an individual that did misreport some of the numbers in the wheel task should feel concerned by the information, except if the biased memory of past decisions provides a wiggle room to persuade himself that he did not cheat that much and that keeping the money is fine. In other words, individuals may bias their recalls downward to distance themselves from dishonest players that are expected to give back part of the money they had earned unethically. In this treatment, minimizing past cheating behavior can have hedonic motives as in the Hedonic treatment (to maintain a moral self-image), but also instrumental motives since unethical forgetting can serve as an excuse for keeping the undeserved money. This leads to our second conjecture:

Conjecture 2 (Instrumental value of memory) *Dishonest individuals exhibit higher positive memory errors when forgetting serves as an excuse to keep undeserved money (Instrumental treatment) than when it only serves hedonic motives (Hedonic treatment).*

3.5 Results

A prerequisite of our experiment is that participants cheat in the wheel task when given the opportunity. Therefore, before presenting the results testing our two conjectures, we first report behavior in part 1 of the experiment.

3.5.1 Cheating Behavior and Classification of Players

To identify individuals who misreported numbers to their advantage, we used a chi-square test on the individuals' distribution of reports. A participant is classified as *dishonest* if the observed frequencies from his reported numbers are significantly different from the expected frequencies of each number (16.67%), at a 10% level.[13] Similarly, a participant is defined as *honest* if his observed frequencies are not statistically different from the expected frequencies. We used the same method to identify participants that had to recall a dishonest versus an honest distribution in the Control treatments.

Overall, 27.01% of participants from the Hedonic and Instrumental treatments were classified as dishonest individuals, and 30.06% of participants from the Hedonic-Control and Instrumental-Control treatments had to recall a dis-

[13]We replicated our analysis using both a stricter definition of a dishonest player (a player is classified as dishonest if his observed and expected frequencies are significantly different at a 5% level), and a less strict definition (a player is classified as dishonest if his observed and expected frequencies are significantly different at a 15% level). The results are robust to any specification (see Table A4.4 in Appendix 4).

honest distribution (see Table A2.1 in Appendix 2). These percentages are not significantly different (two-sample test of proportions, p=0.285). Figure A3.3 in Appendix 3 displays the aggregated distribution of participants' reported numbers, by treatment. None of the distributions is uniform (Kolmogorov-Smirnov one-sample tests, p<0.001). This indicates that our procedure to assign numbers in the Control treatments was effective, allowing us to compare these populations. In all distributions, each number below 4 is significantly less frequently reported than his expected frequency of 16.67% (Student tests, p<0.001) and the frequencies of numbers 5 and 6 are significantly above the expected 16.67% (Student tests, p<0.001). Number 4 is not significantly less frequently reported than his expected frequency of 16.67%, but this is the case regardless of the treatment.

Partial vs. Full Cheaters: Among cheaters, some cheated partially and others cheated to the full extent. We define a full cheater as a participant who reported number 6 in the twenty wheels. Out of 996 participants allocated to the Hedonic and Instrumental treatments, 55 cheated to the full extent (4.10% and 6.89% of full cheaters in the Hedonic and Instrumental treatments, respectively, p=0.053, two-sample test of proportion).[14,15] Because partial and full cheaters may differ in their very nature, the results presented below describe memory errors of dishonest participants both when including full cheaters and when restricting the data to partial cheaters only. The results are robust in both cases. In the last paragraph of this section, we describe memory errors of full cheaters in more detail.

[14]Under no circumstances this difference can be attributed to a treatment effect: part 1 was exactly the same in the Instrumental and Hedonic treatments and participants were assigned one of the two treatments randomly and simultaneously.

[15]This low number of full cheaters may be explained by the participants' desire to balance their self-interest with their moral values. People may cheat to increase their profit but reasonably not to harm their moral self-image (Mazar et al., 2008). These low numbers are also consistent with Gneezy et al. (2018) who find that people report more partial lies when the outcome of a random draw cannot be observed by the experimenter, to preserve their reputation. When the outcome is not observable by the experimenter, a participant who lies may choose to report high but not maximal numbers to signal to the experimenter that he or she does not cheat.

Participants who do not recall the task: Before they were asked to recall the distribution of their past reports, participants had to state whether they remembered the wheel task they performed in part 1. The percentage of dishonest players who stated that they did not remember the wheel task is 14.50% in the Hedonic and Instrumental treatments. In the Control treatments, 11.22% of participants that had to recall a dishonest distribution stated that they did not remember the wheel task. This difference is not significant (p=0.419, two-sample test of proportion). The percentage of players who did not remember the wheel task is also not significantly different between honest individuals and individuals that had to recall a honest distribution (10.73% and 10.53%, respectively, p=0.332). Within treatments, the percentage of players who stated that they did not remember the wheel task is not significantly different between honest and dishonest players (p=0.189 in Hedonic, p=0.316 in Instrumental, p=0.996 in Hedo-Control and p=0.736 in Instru-Control). Therefore, selecting the option "I don't recall" does not seem to depend of the ethical nature of the player (honest *vs.* dishonest).

This is corroborated by a regression analysis presented in Table A2.2 in Appendix 2 that reports Logit regressions in which the dependent variable is equal to one if participants stated that they did not recall the task and zero otherwise. It shows that neither the treatment nor the players' type (honest vs. dishonest) are significantly correlated with the likelihood of stating "I don't recall". The score at the word memory task used to control for individuals' memory cognitive ability, and the age of the participants, however, are strongly correlated (at a 1% level) with the likelihood of reporting "I don't recall". The lower the performance at the word memory task, the higher the likelihood of reporting "I don't recall". Regarding age, the older the participants, the higher the likelihood of reporting "I don't recall". Overall, these findings suggest that participants who selected the option "I

don't recall" did not do it strategically but because they truly did not remember the wheel task they performed in part 1. Since these participants may have a very noisy (if not random) recollection of the numbers they reported in part 1, we restrict the main data analysis to individuals who reported that they remembered the wheel task. The same analysis including participants who reported that they did not remember the wheel task is presented in Table A2.3 in Appendix 2 and does not change qualitatively the results.

3.5.2 Memory of Past Behavior

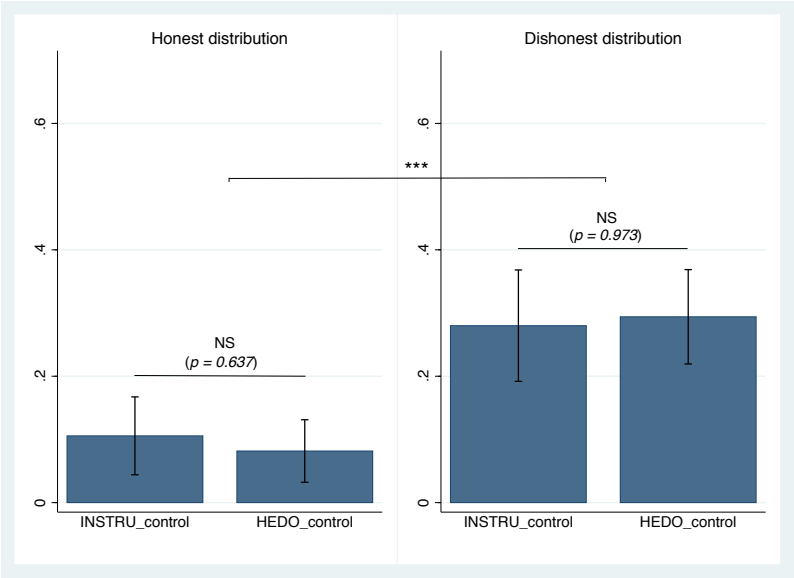
We first analyse the participants' memory errors in the Control treatments. This allows us to measure imperfect memory, i.e., memory errors when there is no intrinsic incentives for memory manipulation. We then present the results from the main treatments in which participants were able to cheat.

Memory Errors in the Control Treatments

We compute the individual's memory errors as the difference between the average reported number in part 1 and the average recalled number in part 2, by individual.[16] Figure 3.2 shows the average memory error in the Hedonic-Control and Instrumental-Control treatments, both for participants that had to recall a honest distribution and participants that had to recall a dishonest one. In the Hedonic-Control treatment, the average memory error of participants that had to recall a dishonest distribution is 0.29. It is 0.28 in the Instrumental-Control treatment. The difference is not significant (p=0.973, M-W test). In the Hedonic-Control treatment, the average memory error of individuals that had to recall a honest distribution is 0.08. It is 0.11 in Instrumental-Control. This difference is not significant either (p=0.637). There is also no difference in memory errors

[16]This measure of memory error is also used in Saucet and Villeval (2019) and Carlson et al. (2018).

(ME hereafter) between the Hedonic-Control and Instrumental-Control treatments when pooling all participants together (ME=0.14 and 0.16 in the Hedonic-Control and Instrumental-Control treatments, respectively, p=0.712, M-W test). Since, as expected, there is no difference in memory errors between the Hedonic-Control and Instrumental-Control treatments, we pooled these two Control treatments called "Pooled-Control" in the data analysis presented below.



Note: The Figure displays the average memory error, by control treatment. P-values are from two-sample Mann-Whitney tests,***1 % level of significance, one observation per individual.

Figure 3.2: Average memory error, Control treatments

Memory Errors in the Main Treatments

We present two results that test our two conjectures. The first result analyzes the existence of motivated memory for hedonic motives. The second result investigates the existence of motivated memory for instrumental motives.

We introduce our first result:

Result 1 (Hedonic value of motivated memory) *Dishonest individuals in the Hedonic treatment do not exhibit higher positive memory errors than individuals that had to recall a dishonest distribution but could not cheat.*

Result 1 rejects Conjecture 1.

Support for Result 1: In the Hedonic treatment, the average memory error of dishonest individuals is 0.31. In the Pooled-Control treatment, the average memory error of participants that had to recall a dishonest distribution is 0.29. The difference is not significant (p=0.354, M-W test). This finding shows that participants who cheated in part 1 did not motivate their memory to make themselves look more honest when recalling their reports in part 2. If this would have been the case, we would have observed higher memory errors from dishonest participants in the Hedonic treatment, in which participants could cheat, than in Pooled-Control treatment in which participants could not cheat and thus had no self-image at stake when recalling.

Table 3.2 reports OLS regressions in which the dependent variable is the average memory error of participants who were classified as dishonest in the main treatments or that had to recall a dishonest distribution in the Pooled-Control treatment. In model (1), the independent variables include the three treatments (with Pooled-Control as the reference category). Model (2) replicates model (1) by controlling for two factors that are predicted to impact memory errors: the standard deviation of participants' reports (since a distribution with a higher variation of reports may be more difficult to recall than a distribution with fewer variation), and participants' score in the word memory task since memory capacity may be heterogeneous across individuals. It also includes the self-reported propensity to

take risks. Model (3) includes four demographic variables: gender, age, monthly expenses and educational attainment. Models (4) to (6) replicate models (1) to (3) but exclude full cheaters. Table 3.2 supports Result 1 on the absence of motivated memory for hedonic reasons. Indeed, participating in the Hedonic treatment does not increase significantly participants' memory errors compared to the Pooled-Control treatment. This is the case regardless of the model specification and of whether full cheaters are included or not. None of the demographic variables (gender, age, monthly expenses and educational attainment) is significantly correlated with memory errors.

Table 3.2: Determinants of memory errors, dishonest players

Dep. var.:	Memory errors					
	WITH FULL CHEATERS			WITHOUT FULL CHEATERS		
	(1)	(2)	(3)	(4)	(5)	(6)
Pooled-Control	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
	-	-	-	-	-	-
HEDO	0.022	0.055	0.060	0.004	-0.010	-0.004
	(0.081)	(0.084)	(0.091)	(0.079)	(0.093)	(0.096)
INSTRU	0.182**	0.221**	0.234**	0.252***	0.237**	0.258**
	(0.089)	(0.094)	(0.103)	(0.097)	(0.093)	(0.102)
Score word task		-0.003	-0.003		-0.002	-0.002
		(0.007)	(0.007)		(0.007)	(0.007)
S.D. report		0.049	0.035		-0.029	-0.041
		(0.073)	(0.074)		(0.123)	(0.121)
Risk		-0.003	0.005		-0.005	0.003
		(0.013)	(0.014)		(0.014)	(0.015)
Male			-0.057			-0.069
			(0.082)			(0.087)
Age			0.006			0.005
			(0.004)			(0.005)
Cons.	0.287***	0.288	0.282	0.287***	0.400	0.404
	(0.058)	(0.228)	(0.327)	(0.058)	(0.301)	(0.396)
Demographics	No	No	Yes	No	No	Yes
N	317	317	317	269	269	269

Note: This Table reports OLS regressions. Robust standard errors clustered at the individual level are in parentheses. Participants who reported "I don't recall" are excluded. One observation per individual. Demographics include monthly expenses and educational attainment. ** $p<0.05$, *** $p<0.01$.

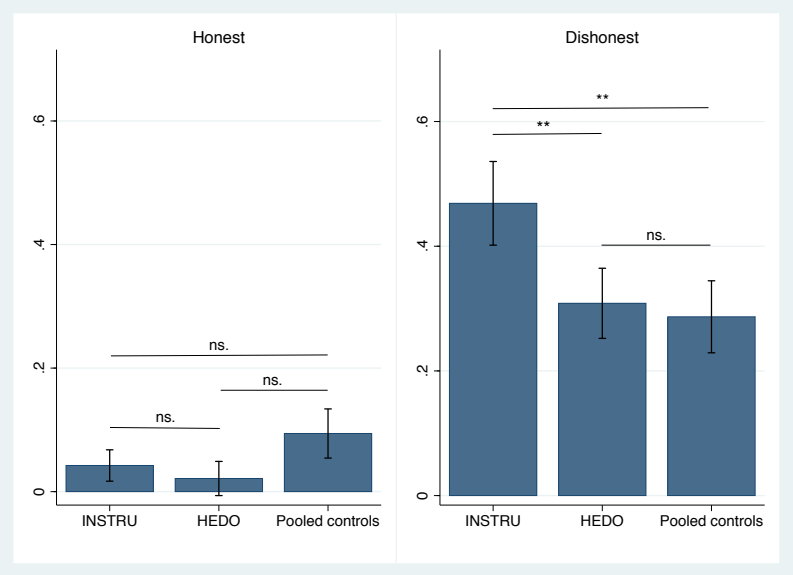
Note that the rejection of Conjecture 1 is very conservative. Indeed, when comparing the Hedonic and Pooled-Control treatments, we are implicitly assuming that participants in both cases paid the same level of attention to the reported numbers in the wheel task. However, in the Hedonic treatment participants had to choose which numbers to report and were thus active in the decision process. This was not the case in the Pooled-Control treatment in which participants had just to report numbers they did not get to choose. As a consequence, participants in the Pooled-Control treatment probably paid less attention to their reports (and encoded them less well) than participants in the main treatments. Thus, inattention could explain a higher share of the memory errors observed in the Pooled-Control treatment than in the Hedonic treatment, resulting in more memory errors in this treatment compared to the Hedonic one.[17]

We now introduce our second result:

Result 2 (Instrumental value of memory) *Dishonest individuals exhibit higher positive memory errors when forgetting is instrumental (i.e., when it serves as a justification to keep undeserved money), than when it only serves hedonic motives (enhancing their self-view). This is not observed for individuals that had to recall a dishonest distribution but could not cheat.*

Result 2 supports Conjecture 2.

[17] To increase attention in the Control treatments, we could have told participants, at the beginning of the wheel task, that they will be asked to recall their reported numbers in the second part taking place three weeks later. However, we decided against it for one main reason: knowing in advance that they will be asked to recall the reports' distribution, participants could have simply written down the reported numbers. Three weeks later, they would just need to report the distribution of numbers they wrote down three weeks before. Since there is no way of checking who took notes and who did not, it would be impossible to know whether the recalled distribution in part 2 captures real memory.



Note: The Figure displays the average memory error of honest and dishonest participants in the main treatments and participants who received honest or dishonest distribution in Pooled-Control treatment. Levels of significance are from Model (1) in Table 3.2 (for dishonest players) and Model (1) in Table A5.5 (for honest players). One observation per individual. Partial cheaters are included. ** p<0.05, ns.= non significant.

Figure 3.3: Average memory error of dishonest players, by treatment

Support for Result 2: Figure 3.3 displays the participants' memory errors across treatments, for both honest and dishonest players. It shows that dishonest participants exhibit higher memory errors in the Instrumental treatment than in the Hedonic and Pooled-Control treatments (at the 5% level). No difference in memory errors is observed between treatments for honest participants (Hedonic and Instrumental treatments) or participants that had to recall a honest distribution (Pooled-Control treatment). If dishonest participants had not manipulated their memory for instrumental reasons, we would have observed no difference in memory errors between the Instrumental and Hedonic treatments, nor between the Instrumental and Pooled-Control treatments.

Table 3.2 supports Result 2 on the existence of memory bias for instrumental reasons. Participating in the Instrumental treatment increases significantly the memory errors of dishonest individuals. This is the case regardless of the model specification and whether full cheaters are included or not. By contrast, participating in the Instrumental treatment does not increase memory errors of honest participants (see Table A5.5 in Appendix 5). Overall, these findings confirm that dishonest individuals forget more their past cheating behavior when it serves as an excuse to keep undeserved money than when it only serves hedonic motives. Only participants who need a justification for keeping the money, i.e., those who have been willfully dishonest, under-estimate their past cheating behavior. If dishonest individuals do not want to give undeserved money back when asked, one strategy is to forget the extent to which they cheated in the first part to convince themselves that they do not belong to the category of dishonest players. On the contrary, individuals that had to recall a dishonest distribution but could not cheat do not need any justification for keeping the money since they are not responsible for the high numbers reported in part 1.

Memory Errors of Partial *vs.* Full Cheaters: The average memory error of full cheaters (those who reported a "6" 20 times) is 0.32. The average memory error of partial cheaters is 0.43. This difference is significant at the 1% level (M-W test, p=0.001). No difference is observed between partial and full cheaters in the word memory task: on average, partial cheaters recalled correctly 25.35 words out of 35 and full cheaters 24.47 (p=0.476, MW). Therefore, the difference in memory errors observed in the wheel game cannot be explained by a difference in cognitive memory ability. Two other reasons may explain this difference in memory errors between partial and full cheaters. First, recalling could be cognitively easier for full cheaters than for partial cheaters who had to recall different numbers.

Models (1) to (6) show that a higher standard deviation of the twenty numbers reported in part 1 does not significantly increase the memory errors in part 2. The correlation between the average memory error and the standard deviation of reports in part 1 is significant neither at the aggregated level (pairwise Pearson's correlation coefficient=-0.015, p=0.792), nor by treatment (pairwise Pearson's correlation coefficient, p=0.858, p=0.674 and p=0.595 in the Pooled-Control, Hedonic and Instrumental treatment, respectively). These results suggest that higher memory errors for partial cheaters are not mainly driven by the fact that a higher variation of reports may be more difficult to recall than a distribution with few variation. Second, the two types of players may have different intrinsic motivations. Full cheaters, who deliberately chose to maximize their payoff in part 1, may feel perfectly fine with this strategy and thus do not need memory manipulation to maintain their moral self-view in part 2. Actually, the fact that none of them decided to give undeserved money back when given this possibility provides some support for this interpretation. If full cheaters accept their greedy nature, they may feel no need to recall selectively or give undeserved money back. By contrast, partial cheaters may have faced a trade-off in part 1 between maximizing their payoff by cheating and maintaining their moral value. Therefore, those participants who cheated to a lesser extent are the ones who use memory manipulation as a self-management strategy.

3.6 Conclusion

This chapter explores the existence of motivated memory in the context of dishonest decision-making. In our experiment, participants played a repeated wheel game allowing them to misreport the outcome of random draws. Three weeks

later, they were asked to recall the distribution of their reports in the previous game. Across treatments, we varied the intrinsic motivation to manipulate one's memory. In the Hedonic treatment, forgetting past misdeeds can have a hedonic value since it may help individuals to preserve a good self-image after misconduct. In the Instrumental treatment, forgetting past misdeeds can additionally have an instrumental value since it may serve as a self-excuse to keep undeserved money. In the Control treatments, both the hedonic and instrumental values of motivated memory are turned off since participants had no possibility to cheat. Overall, this design allowed us to investigate (i) whether individuals manipulated the memory of past dishonest choices, and (ii) whether they used their memory as a self-excuse not to engage in future morally responsible behavior.

We found no clear evidence of motivated memory for purely *hedonic* value. Dishonest individuals in the Hedonic treatment do not exhibit higher memory errors than individuals that had to recall a dishonest distribution in the Control treatments. This does not necessarily contradicts previous evidence. Indeed, this result is very conservative, as memory errors in the Control treatments were possibly overestimated compared to memory errors in the Hedonic treatment. Indeed, in the Hedonic treatment, participants had to choose which numbers to report and were thus active in the decision process. This was not the case in the Control treatments in which participants reported numbers that they did not get to choose. Therefore, memory errors in the Control treatments may capture inattention (due to passiveness), which probably occurs to a lesser extent in the Hedonic treatment in which people were actively dishonest. This is consistent with Saucet and Villeval (2019) who found, using dictator games, that dictators exhibit significantly higher magnitude of memory errors about the receiver's earnings when the later are randomly chosen by the computer program than when they are chosen by the

dictators themselves.

We found evidence of motivated memory for *instrumental* reasons. Dishonest individuals recall their past behavior with less accuracy when they are informed that they will have a future decision to make (whether or not to give undeserved money back) than when they know that they will not have any decision to make. This difference in memory errors is not observed in the Control treatments. Therefore, only participants who need a justification for keeping undeserved money, i.e., those who have been willfully dishonest, underestimate their past cheating behavior. While the previous studies on motivated memory investigate memory manipulation as a *consequence* of past self-image threatening decisions, this result shows that memory manipulation may also occur in reaction to *future* anticipated decisions. This finding also complements the literature on moral wiggle room which shows that individuals exploit uncertainty on the consequences of their actions to behave more selfishly (Dana et al., 2007). In this study, individuals also exploit uncertainty to act more self-interestedly, but uncertainty comes from the imperfect memory of past decisions rather than from the consequences of actions.

Appendices

A1: Instructions

SESSION 1

Overview

If you have any questions regarding this study, you can contact us at saucet@gate.cnrs.fr.

This study is divided into two parts:

- you will complete the first part today;

- you will receive an invitation to complete the second part in three weeks.

For completing today's part, you will receive $1.5. The additional earnings that you will make today will be paid to you at the end of the second part.

IMPORTANT: Please agree to participate in the second part by checking the corresponding box.

◯ I commit to participate in the second part of this study.

The first part consists of two tasks. These two tasks are independent. At the beginning of each task, you will receive the corresponding instructions.

Please press "Next" to continue.

NEXT SCREEN

First task

Choosing a square in a wheel

You will see a wheel with six empty squares. The wheel will look like the one below:

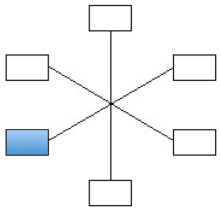


Figure A1.1: Empty Wheel

Your job will be to choose one square of the wheel.

Once you have chosen a square, you will have to press the button "I made my choice".

Please press "Next" to continue.

NEXT SCREEN

Reporting the number displayed in the chosen square

The program will then display a number between 1 and 6 (included) in each square of the wheel. Each square will contain a different number.

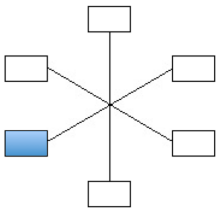
Your job will be to report the number displayed in the square that you have chosen.

Please press "Next" to see an example.

NEXT SCREEN

Example

Before seeing the numbers in the wheel, suppose that you have chosen the square located at the bottom left of the wheel (highlighted in blue in the wheel below for illustration).



You then discover that the numbers are displayed in the wheel as follows:

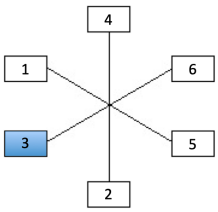


Figure A1.2: Full Wheel

Your job will be to report "3".

Please press "Next" to continue.

NEXT SCREEN

How are your earnings determined?

You will see 20 different wheels one after the other. Before seeing each wheel, you will choose a square, and then, after seeing the numbers in the wheel, you will

report the number corresponding to this square.

At the end of the study, one wheel will be selected at random among the 20. The money your will earn in this task will depend on the number that you have reported for that wheel.

You will earn:

- \$0.10 if you have reported "1"
- \$0.20 if you have reported "2"
- \$0.30 if you have reported "3"
- \$0.40 if you have reported "4"
- \$0.50 if you have reported "5"
- \$0.60 if you have reported "6"

Please press "Next" to continue.

NEXT SCREEN

Method used to choose a square

There are two methods to choose a square in the wheel. Either you will have to choose a square by clicking on it, or you will have to choose a square by keeping it in your mind. We refer to these two methods as the "click" and the "mind" methods.

We have predetermined the proportion of participants who will be assigned to each method and assigned you to one of the two methods when you entered the study.

In the next screen, you will discover which method you have been allocated to. You will keep the same method for the 20 wheels.

Please press "Next" to continue.

NEXT SCREEN

Cheating condition (treatments Hedo and Instru)

You have been allocated to the "mind" method

This means that, for each wheel, you will have to choose one square of the wheel in your mind.

Once you have chosen a square, you will have to press the button "I made my choice".

NEXT SCREEN

No-cheating condition (Hedonic-Control and Instrumental-Control treatment)

You have been allocated to the "click" method

This means that, for each wheel, you will have to choose one square of the wheel by clicking on it. Once you have chosen a square, you will have to press the button "I made my choice".

NEXT SCREEN

End of the instructions for the first task

We would like you not to take any breaks while completing the 20 wheels.

When you are ready, please press "Next" to start the task.

NEXT SCREEN

Cheating condition (treatments Hedo and Instru)

Please choose a square in your mind and then click the button "I made my choice".

No-cheating condition (Hedonic-Control and Instrumental-Control treatments)

Please choose a square and click on it. Then, click the button "I made my choice".

NEXT SCREEN

Please report the number displayed in the square that you have chosen in the previous screen:

NEXT SCREEN

End of the first task

You have completed the wheel task.

Please press "Next" to continue.

NEXT SCREEN

Second task

Words task

This task is totally independent from the first task.

In this task, you will be presented with five lists of words. Each list will contain six words. Each word will be displayed on your screen one by one for less than one second.

Your task consists in memorizing these words.

At the end of today's part, you will be displayed 35 words. For each of these 35 words, your job will be to indicate whether it has occurred on the lists that you have been presented before.

If you remember seeing the word on the lists, you will have to press the button "Old". If you don't remember seeing this word, you will have to press the button "New".

Please press "Next" to watch the five lists of words.

NEXT SCREEN

Participants are displayed with the 5 lists of words.

NEXT SCREEN

End of the lists

You have been displayed the five lists of words.

Please press "Next" to continue.

NEXT SCREEN

Demographic questionnaire

Please answer these few questions about yourself.

NEXT SCREEN

Thank you for completing this short demographic questionnaire.

Recollection of the words

You will now be presented with 35 words. For each of these 35 words, your job will be to indicate whether it has occurred on the lists that you have been presented before.

If you remember seeing the word on the lists, please press the button "Old". If you don't remember seeing this word, please press the button "New".

Please press "Next" to continue.

NEXT SCREEN

How are your earnings determined?

You will be paid $0.02 for each correct recall.

More precisely, you will earn $0.02 per word if:

- You pressed the button "Old" and the word has occurred on the lists.
- You pressed the button "New" and the word has not occurred on the lists.

If you pressed "Old" while the word has not occurred on the lists, or if you pressed "New" while the word has occurred on the lists, you will not earn anything.

When you are ready, please press "Next" to enter your recalls.

NEXT SCREEN

Participants are displayed a list of 35 words. For each of them they have to check either the button "New" or "Old".

NEXT SCREEN

Earnings

Today, you receive $1.5 for having participated in the first part of the study.

In three weeks, you will receive an invitation to participate in the second part. This second part will not exceed 5 minutes.

In the second part of the study, you will receive: $1.5 for participating, plus the earnings that you have made today and the ones that you will make in the second part. If you do not participate in the second part, the earnings that you have made today will be void.

Please press "Next" to continue.

NEXT SCREEN

Thank you!

Thank you for taking time out of your busy life to participate to the first part of this study.

You will receive the $1.5 into your Mturk account within 48 hours.

You will receive an invitation to participate to the second part of this study in three weeks. Once you receive the invitation, you will have three days to complete it.

If you have any questions concerning this study, you can contact us at saucet@gate.cnrs.fr.

Please press "Next" to continue.

NEXT SCREEN

Your confirmation code to be entered on Mturk webpage is your Mturk worker ID.

Please make the HIT on Mturk with this ID.

SESSION 2

Overview

If you have any questions regarding this study, you can contact us at saucet@gate.cnrs.fr.

This is the second and last part of a study on decision-making. You participated in the first part three weeks ago.

Please press "Next" to continue.

NEXT SCREEN

Information

Today, you will earn a fixed payoff of $1.5 for participating to the second part of this study.

At the end of today's part, you may be given or not the option to reduce your total earnings by $0.75.

It would be nice to reduce your total earnings if you are given this option and you have misreported numbers to your advantage several times in the wheel task that you performed three weeks ago.

Remember that in the wheel task, you had to choose a square in a wheel and report the number displayed in this square. You performed that task 20 times.

In the next screen, you will discover whether, at the end of today's part, you will be given or not the option to reduce your total earnings.

Please press "Next" to continue.

NEXT SCREEN

Hedonic condition (Hedonic and Hedonic-Control treatments)

You will not be given the option to reduce your total earnings.

This means that, at the end of today's part, you will not have the option to reduce your total earnings by $0.75.

You will now start the first task of the study.

Please press "Next" to continue.

NEXT SCREEN

Instrumental condition (Instrumental and Instrumental-Control treatments)

You will be given the option to reduce your total earnings.

This means that, at the end of today's part, you will have the option to reduce your total earnings by $0.75.

You will now start the first task of the study.

Please press "Next" to continue.

NEXT SCREEN

Instructions for the first task

We would like you to recall the 20 numbers that you reported in the wheel task you performed three weeks ago. Remember that in this task, you had to choose a square in a wheel and to report the number displayed in this square.

You performed this task 20 times.

Do you remember this task?

YES NO

NEXT SCREEN

You saw 20 wheels one after the other and reported a number after seeing each wheel.

At the end of the study, one wheel was selected at random among the 20. The money you earned in this task depended on the number that you reported for that wheel.

You earned:

- \$0.10 if you reported "1"
- \$0.20 if you reported "2"
- \$0.30 if you reported "3"
- \$0.40 if you reported "4"
- \$0.50 if you reported "5"
- \$0.60 if you reported "6"

Your job consists in recalling how many times you reported the number "1", "2", "3", "4", "5", and "6", respectively, three weeks ago.

Please press "Next" to continue.

NEXT SCREEN

How are your earnings determined?

You will be paid depending on the accuracy of your recalls.

One number ("1", "2", "3", "4", "5", or "6") will be selected at random.

For that number, you will earn:

- \$1 if your recall is accurate (you recall exactly how many times you reported that number three weeks ago).
- \$0.5 if your recall is inaccurate by plus or minus one.
- \$0 if your recall is inaccurate by more than plus or minus one.

Please press "Next" to enter your recalls.

NEXT SCREEN

Three weeks ago, you saw 20 wheels and reported a number for each wheel. Out of 20 wheels:

- How many "1"s did you report?
- How many "2"s did you report?
- How many "3"s did you report?
- How many "4"s did you report?
- How many "5"s did you report?
- How many "6"s did you report?

NEXT SCREEN

End of the task

You have completed the recall task.

Please press "Next" to continue.

NEXT SCREEN

Instrumental condition (Instrumental and Instrumental-Control treatments)

Decision to reduce your total earnings by $0.75

Earlier in the experiment, you have been informed that you will have the option to reduce your total earning by $0.75.

You will now have to decide whether you are willing to reduce or not your total earnings by $0.75.

Please press "Next" to make your choice.

NEXT SCREEN

Instrumental condition (Instrumental and Instrumental-Control treatments)

Are you willing to reduce your total earnings by $0.75 ?

YES NO

NEXT SCREEN

Instrumental condition (Instrumental and Instrumental-Control treatments)

End of the task

You have completed the task.

Please press "Next" to continue.

NEXT SCREEN

(Guilt and religiosity questionnaires)

And finally a few questions about yourself...

NEXT SCREEN

Thank you!

Thank you for taking time out of your busy life to participate to the second and last part of this study.

You will receive the posted remuneration into your Mturk account within 48 hours.

If you have any questions concerning this study, you can contact us at saucet@gate.cnrs.fr.

Please press "Next" to continue.

NEXT SCREEN

Your confirmation code to be entered on Mturk webpage is your Mturk worker ID.

Please make the HIT on Mturk with this ID.

A2: Tables

Table A2.1: Summary Statistics - Participants, by treatment

Treatments	All	Hedo	Instru	Hedo_C	Instru_C
Male	49.84%	48.15%	54.13%	45.40%	46.01%
Age	39.59	40.01	40.22	38.21	37.62
Monthly expenses (0-7 scale)	3.56	3.60	3.58	3.51	3.41
Education (0-4 scale)	2.76	2.79	2.71	2.75	2.82
Religion (0-16 score)	5.23	5.14	4.98	5.75	5.78
% of dishonest players	27.76%	27.25%	26.78%	30.06%	30.06%
% of "I don't recall"	11.50%	11.89%	11.61%	14.11%	07.36%

Notes: This Table reports the results of Pearson Chi-square tests in which each individual is taken as one independent observation. None of the p-values are significant at 5% level or below. The null hypothesis that the coefficients are equal between treatments can thus not be rejected.

Table A2.2: Determinants of "I don't recall"

Dep. var.:	=1 if "I don't recall"					
	WITH FULL CHEATERS			WITHOUT FULL CHEATERS		
	(1)	(2)	(3)	(4)	(5)	(6)
Pooled-Control	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
	-	-	-	-	-	-
HEDO	0.019	0.163	0.0259	0.019	0.166	0.025
	(0.276)	(0.279)	(0.288)	(0.276)	(0.280)	(0.288)
INSTRU	0.024	0.145	-0.019	0.024	0.148	-0.022
	(0.273)	(0.277)	(0.289)	(0.273)	(0.278)	(0.288)
=1 if Dishonest	0.072	0.077	0.104	0.072	0.073	0.103
	(0.386)	(0.393)	(0.406)	(0.386)	(0.394)	(0.407)
=1 if Dishonest * Pooled-Control	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
	-	-	-	-	-	-
=1 if Dishonest * HEDO	0.318	0.374	0.465	0.450	0.455	0.591
	(0.487)	(0.530)	(0.541)	(0.492)	(0.533)	(0.541)
=1 if Dishonest * INSTRU	0.226	0.309	0.282	0.132	0.193	0.149
	(0.488)	(0.535)	(0.546)	(0.515)	(0.556)	(0.568)
Score word task		-0.078***	-0.084***		-0.082***	-0.087***
		(0.019)	(0.019)		(0.020)	(0.020)
S.d. report		0.200	0.055		0.168	0.057
		(0.278)	(0.284)		(0.374)	(0.383)
Risk		0.051	0.059		0.048	0.055
		(0.035)	(0.038)		(0.036)	(0.039)
Age			0.039***			0.039***
			(0.008)			(0.008)
Male			0.385**			0.349*
			(0.189)			(0.193)
Cons.	-2.140***	-0.849	-2.585***	-2.140***	-0.680	-2.497**
	(0.216)	(0.671)	(0.896)	(0.216)	(0.807)	(1.027)
Demographics	No	No	Yes	No	No	Yes
N	1322	1322	1321	1267	1267	1266

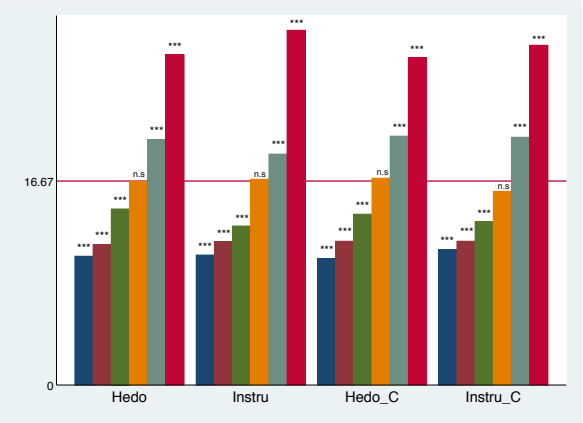
Note: This Table reports results from Logit regressions. Robust standard errors clustered at the individual level are in parentheses. One observation per individual. Demographics: monthly expenses and educational attainment. * p<0.1, ** p<0.05, *** p<0.01.

Table A2.3: Determinants of memory errors, including "I don't recall"

Dep. var.:	Memory errors	
	WITH F.C.	WITHOUT F.C.
	(1)	(2)
Pooled-Control	*ref.*	*ref.*
	-	-
HEDO	0.048	-0.011
	(0.087)	(0.093)
INSTRU	0.180*	0.190*
	(0.099)	(0.098)
Score word task	-0.008	-0.005
	(0.007)	(0.007)
S.D report	0.045	-0.038
	(0.070)	(0.113)
Risk	0.005	0.007
	(0.013)	(0.014)
Male	-0.060	-0.077
	(0.077)	(0.081)
Age	0.004	0.004
	(0.004)	(0.004)
Cons.	0.450	0.471
	(0.324)	(0.361)
Demographics	Yes	Yes
N	367	312

Note: This Table reports OLS regressions. Robust standard errors clustered at the individual level are in parentheses. One observation per individual. F.C. accounts for Full Cheaters. Demographics: monthly expenses and educational attainment. * p<0.10. In model (1), the exact p-value of the INSTRU coefficient is 0.069. In model (2), it is 0.053.

A3: Figure



Note: The Figure displays the percentage of reported number (from 1 to 6) of participants, by treatment. Stars display the significance of two-sided student tests that the observed percentage differs from 16.67%; ***1 % level, one observation per individual.

Figure A3.3: Percentage of reported number of participants, by treatment

A4: Memory Errors with Alternative Definitions of Dishonesty

So far, we identified a dishonest player if the observed frequencies of his reported numbers were significantly different from the expected frequencies at a 10% level. However, the moral cost from remembering dishonest behavior, and thus the intrinsic motivations for memory manipulation, may vary with the very level of dishonesty. The following section uses alternative definitions of dishonesty. Precisely, we consider a stricter definition of a dishonest player by identifying a player as dishonest if his observed and expected frequencies are significantly different at a 5% level. Similarly, we consider a less stric definition of a dishonest player by identifying a player as dishonest if his observed and expected frequencies are significantly different at a 15% level.

Table A4.4 replicates Table 3.2 using these two alternative definitions of dishonesty. First, it shows that participating in the Hedonic treatment does not significantly increase memory errors compared to the Pooled-Control treatment in which participants could not cheat. This is the case regardless of the definition (5% or 15%). Second, it shows that participating in the Instrumental treatment significantly increases the size of memory errors. This is the case regardless of which definition of a dishonest player is being used. Overall, these findings show that Results 1 and 2 are not driven by a specific threshold used to classify players, and are robust to different definitions of a dishonest player.

A5: Memory Errors of Honest Players

Table A5.5 replicates Table 3.2 for honest players. Model (1) shows that participating in the Hedonic treatment or Instrumental treatments does not significantly

Table A4.4: Determinants of memory errors with alternative definitions of dishonesty

Dep. var.:	Memory errors			
	WITH FULL CHEATERS		WITHOUT FULL CHEATERS	
	Disho. 5%	Disho. 15%	Disho. 5%	Disho. 15%
Pooled-Control	*ref.*	*ref.*	*ref.*	*ref.*
	-	-	-	-
HEDO	0.106	0.044	0.025	0.001
	(0.107)	(0.078)	(0.114)	(0.082)
INSTRU	0.285**	0.161*	0.336***	0.180**
	(0.130)	(0.085)	(0.129)	(0.084)
Score word task	-0.001	-0.006	-0.001	-0.006
	(0.007)	(0.006)	(0.008)	(0.007)
S.D. report	0.064	-0.0002	-0.025	-0.078
	(0.084)	(0.068)	(0.134)	(0.113)
Risk	0.0119	0.00272	0.0102	0.0003
	(0.015)	(0.012)	(0.017)	(0.013)
Male	-0.032	-0.046	-0.031	-0.056
	(0.097)	(0.070)	(0.105)	(0.072)
Age	0.005	0.005	0.004	0.004
	(0.005)	(0.003)	(0.005)	(0.004)
Cons.	0.117	0.448	0.288	0.608*
	(0.370)	(0.287)	(0.461)	(0.349)
Demographics	yes	yes	yes	yes
N	247	382	199	334

Notes: This Table reports OLS regressions. Robust standard errors clustered at the individual level are in parentheses. Participants who reported "I don't recall" are excluded. Demographics include monthly expenses and educational attainment. * p<0.1,** p<0.05, *** p<0.01.

increase memory errors compared to the Pooled-Control treatment. Unlike dishonest players, honest individuals had no reason to manipulate their memory since accurate recalls were incentivized. Models (2) and (3) show that participants in the Hedonic treatment exhibit significantly (at a 5% level) lower memory errors than participants that had to recall a honest distribution in the Pooled-Control treatment. This is likely due to the fact than in the Hedonic treatment (as in the Instrumental treatment), participants had to choose which numbers to report and were thus active in the decision process. Models (2) and (3) also show that the standard deviation of the numbers reported in part 1 is significantly (at a 1% level) negatively correlated with memory errors: the higher the standard deviation of reports in part 1, the lower the memory errors in part 2. This might seem counter intuitive since a distribution with a higher variation of reports may be cognitively more difficult to recall than a distribution with few variation. However, for honest

Table A5.5: Determinants of memory errors, honest players

Dep. var.:	Memory errors		
	(1)	(2)	(3)
Pooled-Control	*ref.*	*ref.*	*ref.*
	-	-	-
HEDO	-0.073	-0.111**	-0.113**
	(0.048)	(0.049)	(0.048)
INSTRU	-0.052	-0.081*	-0.070
	(0.047)	(0.047)	(0.048)
Score word task		0.009*	0.008
		(0.005)	(0.005)
S.D. report		-0.416***	-0.402***
		(0.105)	(0.105)
Risk		-0.018***	-0.017**
		(0.007)	(0.007)
Male			-0.026
			(0.036)
Age			0.001
			(0.002)
Cons.	0.094**	0.665***	0.379
	(0.040)	(0.232)	(0.250)
Demographics	no	no	yes
N	853	853	853

Notes: This Table reports marginal effects from an OLS regression. Robust standard errors clustered at the individual level are in parentheses. Participants who reported "I don't recall" are excluded. Demographics include monthly expenses and educational attainment. * p<0.1,** p<0.05, *** p<0.01.

players, a high standard deviation of reports means that they did not cheat in part 1, which is easy for them to remember. Individuals who reported a uniform distribution (characterized by a high standard deviation of reports) are the ones for which recalling is the easiest and thus those who exhibit lower memory errors.

A6: Statistical Power Analysis

To estimate the sample size necessary to uncover the hypothesized effect between the main and the control treatments, we built on the results of Kouchaki and Gino (2016). Their Study 6 is very similar to our design since participants played a die-throwing game and were assigned to one of two conditions: likely-cheating vs. no-cheating. Two days later, they answered questions about their memory of the

die-throwing task on the AMQ measure.[18] In the likely-cheating condition, the average memory accuracy was 4.92 (S.D.=1.16, N=134). In the no-cheating condition, it was 5.54 (S.D.=1.16, N=145). Based on the resulting calculated effect size (Cohen's d=0.5652), and assuming a type-I error rate of 0.10 and a power level of 0.8, the required sample size is 40 observations per treatment. Thus, we are relatively confident that we can uncover the hypothesized effect between the main and control treatments with our sample size.

For the main treatments, there was no existing study that could give us an insight on the expected effect size between the Hedonic and Instrumental conditions. Thus, we determined our sample size using the cohen's d small-medium standard value (0.35), assuming a type-I error rate of 0.10 and a power level of 0.8. With this effect size, the required number of observations per treatment is 102. Here again, we can be relatively confident that the hypothesized effect between the Hedonic and Instrumental treatment can be uncovered since we have 133 and 136 dishonest participants in the Hedonic and Instrumental treatment, respectively.

[18]The AMQ measures self-reported people's autobiographical memory with several items (e.g., "As I think about the coin-toss task/dinner that night, I can actually remember it"), see Rubin et al. (2003).

General Conclusion

This thesis investigates the existence and strength of memory distortion in three economically relevant contexts: social preferences, individual performance and unethical behavior. It explores two different underlying mechanisms of motivated memory: hedonic when memory manipulation only makes oneself look better in one's own eyes, or instrumental when it also helps justify future decisions.

Chapter one investigates whether people retrieve their memory self-servingly in social encounters. Our results identify a causal effect of the responsibility of pro-social decisions on selective recalls. When individuals are responsible for the outcome of a sharing decision, they exhibit higher memory accuracy of altruistic decisions compared to selfish decisions. This is not the case when the outcome of the sharing-decision is randomly chosen by the computer. This result shows that individuals exhibit selective memory of their pro-sociality. Second, incentivizing correct recalls increases correct recalls but only when participants made an altruistic decision. This suggests that, when incentivized, people allocate extra memory effort to retrieve the memory of desirable rather than undesirable information. Finally, selective recalls are not driven by a higher attention paid to the other's amount by other-regarding individuals since individuals are also less likely to remember their own amount after they made a selfish than an altruistic decision. In contrast, we find no clear evidence of motivated memory through biased memory

errors. Individuals do not exhibit overly optimistic recalls of the amounts allocated to the other players. There are two main explanations for the absence of biased memory errors. First, even if the majority of individuals probably prefer to think of themselves as generous rather than egoistic, making selfish decisions –especially in the lab– may not be threatening enough to trigger self-serving memory errors. This might be particularly the case in our context where only self-image, not social-image, is at stake. Uncovering how memory for selfish *vs.* egoistic decisions shifts as a function of direct observability by others would be informative. Investigating motivated memory in the domain of morality and ethics, where categorical imperative (Kant, 1785) and injunctive norms are more salient, could also generate a stronger need for biased memory errors. Finally, in our study the only reason for memory manipulation was hedonic (to make oneself look more generous). An interesting extension (investigated in Chapter 3 about unethical behavior) would be to introduce strategic reasons to test whether individuals would use selective memory and bias their recalls asymmetrically.

Chapter two disentangles between two driving forces that have been proposed as explanations of memory failures for self-relevant information: self-enhancement and mood-congruency. The first result of this chapter show that, consistent with self-enhancing memory, individuals exhibit higher memory accuracy of positive than negative feedback about their cognitive ability. In contrast, we do not find clear evidence of mood-congruent memory. Indeed, even though our mood-manipulation proves to be effective in inducing the desired emotional state, individuals do not exhibit higher memory accuracy when the feedback to retrieve is congruent to their mood. Taken together, these results provide support for the existence and relative dominance of self-enhancing memory over mood-congruent memory. Thereby, it underlines the importance of motivational over affective fac-

tors in the formation of optimistic beliefs about the self. Insofar individuals are able to forget or distort feedback that challenges their beliefs, providing them with self-relevant information may thus not be the most efficient policy to help mitigate or remove biased judgments. One potential explanation of why, in our setting, mood cannot be identified as a driving force of feedback retrieval, is that the similarity between the valence of the mood and the valence of feedback was to weak to trigger associative memory and thereby "compete" against the self-enhancement hypothesis. In real life, self-relevant feedback can be embedded with negative emotions but with also other types of cues such that situations, images, sounds, etc. (Enke et al., 2019). Therefore, an alternative possibility to investigate the relative role of associative memory and self-enhancing memory in asymmetric recall of feedback would be to use such other types of context-dependent and not mood-dependent cues.

Chapter three investigates the relative role of affect and strategic reasoning in motivated memory, with an application in the domain of unethical behavior. We find that hedonic considerations, in our setting, are not sufficient to trigger memory manipulation: dishonest individuals do not exhibit lower memory accuracy of their past decisions than individuals in the control treatment. One limit of our design is that participants in the main treatment were more active in the decision process than participants in the control treatment. Therefore, memory errors in the control treatment may capture inattention (due to passiveness), which may hide a memory bias for hedonic reasons in the treatment where people could cheat. It would be interesting to see if this result holds in a situation where the hedonic reasons to motivate one's memory remain absent for participants in the control treatment but all participants are similarly active in the decision process. In contrast, we have a remarkable result in the Instrumental treatment. When

forgetting serves as a justification to *not* engage in *future* morally responsible behavior, individuals do motivate their memory. Indeed, only participants who need a justification for keeping undeserved money, i.e., those who have been dishonest, under-estimate their past cheating behavior. This result suggests that memory may enter into the decision process not only as a reaction to *past* decisions but also in reaction to anticipated *future* decisions.

Overall, our results show that memory distortions can result from cognitive impairment (imperfect memory) but also from motivated memory. First, individuals manipulate their memory for hedonic reasons to sustain their demand for positive self-image, whether "positive" means generous, honest and/or intelligent. Second, individuals manipulate their memory for strategic reasons. They exploit uncertainty from imperfect memory of past decisions to justify acting self-interestedly in anticipated future decisions. Such memory distortion in reaction to *anticipated* decisions echoes the very recent literature on the "memory of the future" that has developed the past few years in neurosciences to examine the role of memory in *future* thinking (Schacter et al., 2012; Eustache, 2019). Memory, which by its very nature seems to be oriented towards the *past*, would in fact be also intrinsically and deliberately oriented towards the *future*. Not only individuals would be able to retrieve past actions, but they would also store and recall the actions they program or plan for the future. Those "memories of the future" would form the basis for anticipations and expectations and would be continuously rehearsed and optimized for decision-making. Economists may be willing to explore this uncovered aspect of humans' memory.

Bibliography

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*.

Akerlof, G. A. and Dickens, W. T. (1982). The economic consequences of cognitive dissonance. *The American economic review*, 72(3):307–319.

Andrade, E. B., Odean, T., and Lin, S. (2015). Bubbling with excitement: an experiment. *Review of Finance*, 20(2):447–466.

Andreoni, J., Nikiforakis, N., and Stoop, J. (2017). Are the rich more selfish than the poor, or do they just have more money? a natural field experiment. *Working paper, University of California San Diego, San Diego*.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1):544–555.

Azar, O. H., Yosef, S., and Bar-Eli, M. (2013). Do customers return excessive change in a restaurant?: A field experiment on dishonesty. *Journal of Economic Behavior & Organization*, 93:219–226.

Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). Biased judgments of fairness in bargaining. *The American Economic Review*, 85(5):1337–1343.

Baddeley, A. D. (1997). *Human memory: Theory and practice*. Psychology Press.

Banaji, M. R., Bazerman, M. H., and Chugh, D. (2004). How (un)ethical are you? *Revista Icade. Revista de las Facultades de Derecho y Ciencias Económicas y Empresariales*, (62):359–365.

Banaji, M. R. and Bhaskar, R. (2000). *Implicit stereotypes and memory: The bounded rationality of social beliefs*. Cambridge, Massachusetts: Harvard University Press.

Bartlett, F. C. (1932). Remembering: An experimental and social study. *Cambridge University Press*.

Bartling, B. and Fischbacher, U. (2011). Shifting the blame: On delegation and responsibility. *The Review of Economic Studies*, 79(1):67–87.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *American Economic Review*, 97(2):170–176.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, 5(4):323–370.

Bazerman, M. H., Tenbrunsel, A. E., and Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Making decisions with competing internal preferences. *Academy of Management Review*, 23(2):225–241.

Bechara, A., Damasio, H., Tranel, D., and Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of neuroscience*, 18(1):428–437.

Bénabou, R. (2015). The economics of motivated beliefs. *Revue d'économie politique*, 125(5):665–685.

Bénabou, R., Falk, A., and Tirole, J. (2018). Narratives, imperatives and moral reasoning. *CEPR Discussion Paper, 13056.*

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Bénabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy*, 112(4):848–886.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Bénabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855.

Bernheim, B. D. and Thomadsen, R. (2005). Memory and anticipation. *The Economic Journal*, 115(503):271–304.

Bicchieri, C., Dimant, E., and Sonderegger, S. (2019). It's not a lie if you believe it: Lying and belief distortion under norm-uncertainty. *Available at SSRN.*

Blanchflower, D. G. and Oswald, A. J. (2008). Is well-being u-shaped over the life cycle? *Social science & medicine*, 66(8):1733–1749.

Blaney, P. H. (1986). Affect and memory: a review. *Psychological bulletin*, 99(2):229.

Bless, H., Fiedler, K., and Forgas, J. (2006). Mood and the regulation of information processing and behavior. *Affect in social thinking and behavior*, 6584.

Blustein, J. (2017). A duty to remember. In *The Routledge Handbook of Philosophy of Memory*, pages 351–363. Routledge.

Bock, O., Baetge, I., and Nicklisch, A. (2014). Hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.

Bodoh-Creed, A. (2017). Mood, memory, and biased beliefs and decisions.

Bordalo, P., Coffman, K., Gennaioli, N., Schwerter, F., and Shleifer, A. (2019). Memory and representativeness.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2017). Memory, attention, and choice. *NBER Working Paper No. 23256*.

Bourget, D. and Bradford, J. M. (1995). Sex offenders who claim amnesia for their alleged offense. *Bulletin of the American Academy of Psychiatry & the Law*.

Bower, G. H. (1981). Mood and memory. *American psychologist*, 36(2):129.

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.

Breaban, A., Van De Kuilen, G., and Noussair, C. N. (2016). Prudence, emotional state, personality, and cognitive ability. *Frontiers in psychology*, 7:1688.

Bruhin, A., Fehr, E., and Schunk, D. (2018). The many faces of human sociality: Uncovering the distribution and stability of social preferences. *Journal of the European Economic Association, Forthcoming*.

Brunnermeier, M. K. and Parker, J. A. (2005). Optimal expectations. *American Economic Review*, 95:1092–1118.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American economic review*, 89(1):306–318.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.

Capra, M. C. (2004). Mood-driven behavior in strategic interactions. *American Economic Review*, 94(2):367–372.

Carlson, R. W., Marechal, M., Oud, B., Fehr, E., and Crockett, M. (2018). Motivated misremembering: Selfish decisions are more generous in hindsight. *Mimeo, https://doi.org/10.31234/osf.io/7ck25*.

Carrillo, J. D. and Mariotti, T. (2000). Strategic ignorance as a self-disciplining device. *Review of Economic Studies*, 67(3):529–544.

Casari, M. and Cason, T. N. (2009). The strategy method lowers measured trustworthy behavior. *Economics Letters*, 103(3):157–159.

Chew, S. H., Huang, W., and Zhao, X. (2018). Motivated false memory. *Mimeo, National University of Singapore, February*, page http://dx.doi.org/10.2139/ssrn.2127795.

Chrobak, Q. M. and Zaragoza, M. S. (2008). Inventing stories: Forcing witnesses to fabricate entire fictitious events leads to freely reported false memories. *Psychonomic Bulletin & Review*, 15(6):1190–1195.

Chugh, D., Bazerman, M. H., and Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. *Conflicts of interest: Challenges and solutions in business, law, medicine, and public policy*, pages 74–95.

Cima, M. J., Merckelbach, H., Nijman, H., Knauer, E., and Hollnack, S. (2002). I can't remember your honor: Offenders who claim amnesia. *The German Journal of*

Psychiatry, 5:24–34.

Cohn, A., Maréchal, M. A., Tannenbaum, D., and Zünd, C. L. (2019). Civic honesty around the globe. *Science*, page eaau8712.

Cojoc, D. and Stoian, A. (2014). Dishonesty and charitable behavior. *Experimental Economics*, 17(4):717–732.

Compte, O. and Postlewaite, A. (2004). Confidence-enhanced performance. *American Economic Review*, 94(5):1536–1557.

Conen, M., Cecchin, G., and Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Crary, W. G. (1966). Reactions to incongruent self-experiences. *Journal of Consulting Psychology*, 30(3):246.

Dai, Z., Galeotti, F., and Villeval, M. C. (2017). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, 64(3):1081–1100.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Deffains, B., Espinosa, R., and Thöni, C. (2016). Political self-serving bias and redistribution. *Journal of Public Economics*, 134:67–74.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The quarterly journal of economics*, 127(1):1–56.

Dessi, R., Gallo, E., and Goyal, S. (2016). Network cognition. *Journal of Economic Behavior & Organization*, 123:78–96.

Dow, J. (1991). Search decisions with limited memory. *Review of Economic Studies*, 58(1):1–14.

Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in disguise–a theoretical analysis of cheating. *Journal of Economic Theory*, 175:248–264.

Edmans, A., Garcia, D., and Norli, Ø. (2007). Sports sentiment and stock returns. *The Journal of Finance*, 62(4):1967–1998.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–38.

Enke, B., Schwerter, F., and Zimmermann, F. (2019). Associative memory and belief formation. *Working paper*.

Epstein, S. (1973). The self-concept revisited: Or a theory of a theory. *American Psychologist*, 28(5):404.

Eustache, F. (2019). *La mémoire au futur*. Le Pommier.

Exley, C. L. (2015). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies*, 83(2):587–628.

Exley, C. L. and Kessler, J. B. (2018). Motivated cognitive limitations. *Mimeo, Harvard Business School*.

Feiler, L. (2014). Testing models of information avoidance with binary choice dictator games. *Journal of Economic Psychology*, 45:253–267.

Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation.

Fiedler, K. and Hütter, M. (2013). Memory and emotion. *Sage handbook of applied memory*, pages 145–161.

Fiedler, K., Nickel, S., Muehlfriedel, T., and Unkelbach, C. (2001). Is mood congruency an effect of genuine memory or response bias? *Journal of Experimental social psychology*, 37(3):201–214.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance*, 3(4):552.

Foerster, M. and van der Weele, J. J. (2018). Denial and alarmism in collective action problems.

Foerster, M. and Van der Weele, J. J. (2018). Persuasion, justification and the communication of social impact.

Franzen, A. and Pointner, S. (2013). The external validity of giving in the dictator game. *Experimental Economics*, 16(2):155–169.

Garcia, T., Massoni, S., and Villeval, M. C. (2018). Ambiguity and excuse-driven behavior in charitable giving. *Available at SSRN 3283773*.

Gervais, S. and Odean, T. (2001). Learning to be overconfident. *The Review of Financial Studies*, 14(1):1–27.

Gilovich, T. (2008). *How we know what isn't so*. Simon and Schuster.

Gneezy, U., Imas, A., and Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science*, 60(11):2645–2658.

Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–53.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Gödker, K., Jiao, P., and Smeets, P. (2019). Investor memory. *Available at SSRN 3348315*.

Golman, R., Hagmann, D., and Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, 55(1):96–135.

Gonsalves, B. and Paller, K. A. (2002). Mistaken memories: remembering events that never happened. *The Neuroscientist*, 8(5):391–395.

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M.-M., and Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science*, 15(10):655–660.

Gottlieb, D. (2014). Imperfect memory and choice under risk. *Games and Economic Behavior*, 85:127–158.

Green, J. D. and Sedikides, C. (2004). Retrieval selectivity in the processing of self-referent information: Testing the boundaries of self-protection. *Self and Identity*, 3(1):69–80.

Green, J. D., Sedikides, C., Pinter, B., and Van Tongeren, D. R. (2009). Two sides to self-protection: Self-improvement strivings and feedback from close relationships eliminate mnemic neglect. *Self and Identity*, 8(2-3):233–250.

Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35(7):603.

Grossman, Z. (2014). Strategic ignorance and the robustness of social preferences. *Management Science*, 60(11):2659–2665.

Grossman, Z. and Van Der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1):173–217.

Hammond, J. S., Keeney, R. L., and Raiffa, H. (2006). The hidden traps in decision making. *Harvard Business Review*, 84(1):118.

Hilton, H. (2006). Quelques aspects de la mémoire verbale en l2. *Recherche et pratiques pédagogiques en langues de spécialité. Cahiers de l'Apliut*, 25(2):44–60.

Hirshleifer, D. and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032.

Iriberri, N. and Rey-Biel, P. (2011). The role of role uncertainty in modified dictator games. *Experimental Economics*, 14(2):160–180.

Irlenbusch, B. and Villeval, M. C. (2015). Behavioral ethics: how psychology influenced economics and how economics might inform psychology? *Current Opinion in Psychology*, 6:87–92.

Isen, A. M., Shalker, T. E., Clark, M., and Karp, L. (1978). Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of personality and social psychology*, 36(1):1.

Jacobsen, C., Fosgaard, T. R., and Pascual-Ezama, D. (2018). Why do we lie? a practical guide to the dishonesty literature. *Journal of Economic Surveys*, 32(2):357–387.

Jiang, T. (2013). Cheating in mind games: The subtlety of rules matters. *Journal of Economic Behavior & Organization*, 93:328–336.

Kahana, M. J. (2012). *Foundations of human memory*. OUP USA.

Kajackaite, A. (2015). If i close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment. *Journal of Economic Behavior & Organization*, 116:518–524.

Kajackaite, A. (2018). Lying about luck versus lying about performance. *Journal of Economic Behavior & Organization*, 153:194–199.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2003). Winter blues: A sad stock market cycle. *American Economic Review*, 93(1):324–343.

Kant, E. (1785). *Fondements de la métaphysique des mœurs (1785)*.

Karlsson, N., Loewenstein, G., and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and uncertainty*, 38(2):95–115.

Keizer, K., Lindenberg, S., and Steg, L. (2008). The spreading of disorder. *Science*, 322(5908):1681–1685.

Khalmetski, K. and Sliwka, D. (2019). Disguising lies-image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics, forthcoming*.

Kirchsteiger, G., Rigotti, L., and Rustichini, A. (2006). Your morals might be your moods. *Journal of Economic Behavior & Organization*, 59(2):155–172.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *American Economic Review*, 90(4):1072–1091.

Kopelman, M. D., Green, R., Green, E., Lewis, P., and Stanhope, N. (1994). The case of the amnesic intelligence officer. *Psychological medicine*, 24(4):1037–1045.

Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human learning and memory*, 6(2):107.

Korner, I. N. (1950). *Experimental Investigation of some aspects of the problem of repression: repressive forgetting*. Number 970. Bureau of Publications, Teachers College, Columbia University.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4(4):673–707.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Kouchaki, M. and Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences*, 113(22):6166–6171.

Kuiper, N. A. and Derry, P. A. (1982). Depressed and nondepressed content self-reference in mild depressives. *Journal of personality*, 50(1):67–80.

Li, K. K. (2013). Asymmetric memory recall of positive and negative events in social interactions. *Experimental Economics*, 16(3):248–262.

Li, K. K. (2017). What determines overconfidence and memory recall bias? the role of feedback, awareness and social comparison. *Mimeo, City University of Hong Kong*.

Lykken, D. and Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological science*, 7(3):186–189.

Malmendier, U. and Tate, G. (2005). Ceo overconfidence and corporate investment. *The journal of finance*, 60(6):2661–2700.

Markus, H. and Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, 38(1):299–337.

Martin, J. M., Lejarraga, T., and Gonzalez, C. (2018). The effects of motivation and memory on the weighting of reference prices. *Journal of Economic Psychology*, 65:16–25.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*, 45(6):633–644.

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences*, 20(1):91–102.

Mijović-Prelec, D. and Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1538):227–240.

Mill, J. S. (1874). *Essays on some unsettled questions of political economy*. JW Parker.

Mischel, W. (1979). On the interface of cognition and personality: Beyond the person–situation debate. *American Psychologist*, 34(9):740.

Mischel, W., Ebbesen, E. B., and Zeiss, A. M. (1976). Determinants of selective memory about the self. *Journal of Consulting and Clinical Psychology*, 44(1):92.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence. Technical report, National Bureau of Economic Research.

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

Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*, 117(3):735–774.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5):482.

Naka, M. and Naoi, H. (1995). The effect of repeated writing on memory. *Memory & cognition*, 23(2):201–212.

Nguyen, Y. and Noussair, C. N. (2014). Risk aversion and emotions. *Pacific economic review*, 19(3):296–312.

Nietzsche (1886). *Beyond good and evil*. Boni & Liveright, Incorporated.

Nisan, M. and Kurtines, W. (2013). The moral balance model: Theory and research extending our understanding of moral choice and deviation. *Handbook of Moral Behavior and Development Application*, pages 213–249.

O'Connor, K. M., De Dreu, C. K., Schroth, H., Barry, B., Lituchy, T. R., and Bazerman, M. H. (2002). What we want to do versus what we think we should do: An empirical investigation of intrapersonal conflict. *Journal of Behavioral Decision Making*, 15(5):403–418.

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of finance*, 53(5):1775–1798.

Oexl, R. and Grossman, Z. J. (2013). Shifting the blame to a powerless intermediary. *Experimental Economics*, 16(3):306–312.

Oster, E., Shoulson, I., and Dorsey, E. (2013). Optimal expectations and limited medical testing: evidence from huntington disease. *American Economic Review*, 103(2):804–30.

Oswald, A. J., Proto, E., and Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics*, 33(4):789–822.

Piccione, M. and Rubinstein, A. (1997). On the interpretation of decision problems with imperfect recall. *Games and Economic Behavior*, 20(1):3–24.

Ploner, M. and Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, 93:374–383.

Potters, J. and Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review*, 87:26–33.

Rabin, M. and Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The quarterly journal of economics*, 114(1):37–82.

Raslau, F., Klein, A., Ulmer, J., Mathews, V., and Mark, L. (2014). Memory part 1: Overview. *American Journal of Neuroradiology*, 35(11):2058–2060.

Roediger, H. L. and McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of experimental psychology: Learning, Memory, and Cognition*, 21(4):803.

Rosenbaum, S. M., Billinger, S., and Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, 45:181–196.

Rubin, D. C., Schrauf, R. W., and Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & cognition*, 31(6):887–901.

Saucet, C. and Villeval, M. C. (2019). Motivated memory in dictator games. *Games and Economic Behavior, forthcoming*.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of experimental psychology: learning, memory, and cognition*, 13(3):501.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., and Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, 76(4):677–694.

Schaefer, A., Nils, F., Sanchez, X., and Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion*, 24(7):1153–1172.

Schelling, T. C. (1987). The mind as a consuming organ. *The multiple self*, pages 177–96.

Schwarz, N. (1988). How do i feel about it? the informative function of mood. *Affect, cognition and social behavior*.

Schwarz, N. and Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of personality and social psychology*, 45(3):513.

Schwarz, N. and Clore, G. L. (1996). Feelings and phenomenal experiences. *Social psychology: Handbook of basic principles*, 2:385–407.

Schwarzer, R., Jerusalem, M., et al. (1995). Generalized self-efficacy scale. *Measures in health psychology: A user's portfolio. Causal and control beliefs*, 1(1):35–37.

Sedikides, C. and Green, J. D. (2000). On the self-protective nature of inconsistency-negativity management: Using the person memory paradigm to examine self-referent memory. *Journal of personality and social psychology*, 79(6):906.

Sedikides, C. and Green, J. D. (2004). What i don't recall can't hurt me: Information negativity versus information inconsistency as determinants of memorial self-defense. *Social cognition*, 22(1: Special issue):4–29.

Sedikides, C. and Green, J. D. (2009). Memory as a self-protective mechanism. *Social and Personality Psychology Compass*, 3(6):1055–1068.

Shalvi, S., Dana, J., Handgraaf, M. J., and De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2):181–190.

Shalvi, S. and De Dreu, C. K. (2014). Oxytocin promotes group-serving dishonesty. *Proceedings of the National Academy of Sciences*, 111(15):5503–5507.

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2):125–130.

Shalvi, S., Soraperra, I., van der Weele, J. J., and Villeval, M. C. (2019). Shooting the messenger? supply and demand in markets for willful ignorance.

Shu, L. L., Gino, F., and Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37(3):330–349.

Singer, J. A. and Salovey, P. (1996). Motivated memory: Self-defining memories, goals, and affect regulation. *Striving and feeling: Interactions among goals, affect, and self-regulation*, pages 229–250.

Skala, D. (2008). Overconfidence in psychology and finance-an interdisciplinary literature review. *Bank I kredyt*, (4):33–50.

Skinner, C. H., McLaughlin, T., and Logan, P. (1997). Cover, copy, and compare: A self-managed academic intervention effective across skills, students, and settings. *Journal of Behavioral Education*, 7(3):295–306.

Stanley, M. L., Henne, P., Iyengar, V., Sinnott-Armstrong, W., and De Brigard, F. (2017). I'm not the person i used to be: The self and autobiographical memories of immoral actions. *Journal of Experimental Psychology: General*, 146(6):884.

Story, A. L. (1998). Self-esteem and memory for favorable and unfavorable personality feedback. *Personality and Social Psychology Bulletin*, 24(1):51–64.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta psychologica*, 47(2):143–148.

Swihart, G., Yuille, J., and Porter, S. (1999). The role of state-dependent memory in "red-outs". *International Journal of Law and Psychiatry*, 22(3-4):199–212.

Taylor, P. J. and Kopelman, M. D. (1984). Amnesia for criminal offences. *Psychological medicine*, 14(3):581–588.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological bulletin*, 110(1):67.

Taylor, S. E. and Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, 103(2):193.

Tenbrunsel, A. E., Diekmann, K. A., Wade-Benzoni, K. A., and Bazerman, M. H. (2010). The ethical mirage: A temporal explanation as to why we are not as ethical as we think we are. *Research in Organizational Behavior*, 30:153–173.

Thompson, L. and Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, 51(2):176–197.

Tranel, D., Damasio, A. R., Damasio, H., and Brandt, J. P. (1994). Sensorimotor skill learning in amnesia: additional evidence for the neural basis of nondeclarative memory. *Learning & Memory*, 1(3):165–179.

Trivers, R. (2011). *Deceit and self-deception: Fooling yourself the better to fool others*. Penguin UK.

Tulving, E. et al. (1972). Episodic and semantic memory. *Organization of Memory*, 1:381–403.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1-2):231–270.

Van Landeghem, B. (2012). A test for the convexity of human well-being over the life cycle: Longitudinal evidence from a 20-year panel. *Journal of Economic Behavior & Organization*, 81(2):571–582.

Von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung*, 18(1):299–342.

Williams, E. F. and Gilovich, T. (2008). Do people really believe they are above average? *Journal of Experimental Social Psychology*, 44(4):1121–1128.

Wilson, A. (2014). Bounded memory and biases in information processing. *Econometrica*, 82(6):2257–2294.

Woods, D. and Servátka, M. (2019). Nice to you, nicer to me: Does self-serving generosity diminish the reciprocal response? *Experimental Economics*, 22(2):506–529.

Zimmermann, F. (2019). The dynamics of motivated beliefs. *American Economic Review, forthcoming*.

Essais en économie comportementale sur la mémoire motivée

Charlotte Saucet

Résumé

Cette thèse cherche à déterminer si les individus manipulent leur mémoire pour oublier certaines informations qui menacent leurs croyances. Elle teste expérimentalement l'existence et la force de la mémoire motivée dans trois contextes économiquement pertinents : les préférences sociales, la performance individuelle et les décisions malhonnêtes.

Le chapitre 1 examine si les individus font preuve de mémoire motivée dans les interactions sociales. Les individus oublient-ils les conséquences de leurs actes sur autrui ? Le cas échéant, cela dépend-il de la nature (par exemple, égoïste ou altruiste) de leurs actes ? Nos résultats confirment la sélectivité des souvenirs. Les individus se souviennent mieux des conséquences de leurs actions sur autrui lorsqu'ils ont été généreux que lorsqu'ils ont été égoïstes. En revanche, la direction et la magnitude des erreurs de mémoire ne diffèrent pas selon la nature des choix.

Le chapitre 2 démêle deux mécanismes identifiés comme explications possibles de l'existence de mémoire sélective concernant les performances individuelles : l'auto-renforcement et la congruence de l'humeur. Nous testons l'existence de mémoire motivée dans un environnement contrôlé où les deux théories offrent des prédictions divergentes. Nos résultats supportent l'existence et la dominance relative de l'effet d'auto-renforcement par rapport à la congruence de l'humeur, et soulignent ainsi l'importance des facteurs motivationnels plutôt qu'affectifs dans la formation de croyances motivées.

Le chapitre 3 examine si les individus oublient leurs comportements malhonnêtes, non seulement pour des motifs hédoniques mais aussi pour des raisons stratégiques, lorsque l'oubli sert à justifier une décision *future*. Nous trouvons que les considérations hédoniques seules ne sont pas suffisantes pour déclencher une manipulation de la mémoire. En revanche, lorsqu'oublier sert d'excuse pour ne pas avoir à s'engager dans un comportement moralement responsable, les individus manipulent leur mémoire.

Ces résultats montrent que les erreurs de mémoire dans les contextes économiques peuvent résulter d'une déficience cognitive mais aussi d'une mémoire motivée par la volonté de ne pas avoir à se confronter à des informations pouvant nuire à l'image de soi et remettre en cause ses choix futurs.

Mots-Clés: Mémoire sélective, Croyances motivées, Oubli, Expériences

Essays on the behavioral economics of motivated memory

Charlotte Saucet

Abstract

This thesis investigates whether individuals use their memory as a self-deceptive strategy to sustain motivated beliefs. It tests experimentally the existence and strength of memory manipulation in three economically relevant contexts: social interactions, individual performance and unethical decisions.

Chapter one investigates whether people retrieve their memory self-servingly in social encounters. Do individuals forget the consequences of their actions on others? If so, does it depend on the nature (e.g. selfish or altruistic) of the action? Our results identify a causal effect of the responsibility of pro-social decisions on selective recalls. In contrast, there is no clear evidence of biased memory errors.

Chapter two disentangles between two driving forces that have been proposed as explanations of memory failures for self-relevant information: self-enhancement and mood-congruency. We provide a controlled environment where the two theories predict different outcomes. Our results provide support for the existence and relative dominance of self-enhancing memory over mood-congruent memory and thereby underline the importance of motivational factors in the formation of optimistic beliefs about the self.

Chapter three investigates the relative role of affect and strategic reasoning in motivated memory, with an application in the domain of unethical behavior. We study whether individuals manipulate the memory of past dishonest choices, and whether they use their memory as an instrument to justify future decisions. We find that hedonic considerations are not sufficient to trigger memory manipulation. When forgetting serves as a justification to *not* engage in *future* morally responsible behavior, however, individuals do motivate their memory.

Together, these results show that memory errors in economic contexts can result from cognitive impairment but also from memory distortion motivated by the willingness to protect one's self-image and future choices.

Keywords: Selective memory, Motivated beliefs, Forgetting, Experiment