



HAL
open science

Three essays on using Nudges in business firms

Matthieu Plonquet

► **To cite this version:**

Matthieu Plonquet. Three essays on using Nudges in business firms. Economics and Finance. Université Panthéon-Sorbonne - Paris I, 2019. English. NNT : 2019PA01E062 . tel-02524752

HAL Id: tel-02524752

<https://theses.hal.science/tel-02524752>

Submitted on 30 Mar 2020

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

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



Université Paris 1 – Panthéon Sorbonne
UFR d'Économie

Laboratoire de rattachement : CES

Pour l'obtention du titre de Docteur en Économie
Présenté et soutenu publiquement le 17 décembre 2019 par
Matthieu PLONQUET

Three Essays on Using Nudges in Business Firms

Sous la supervision de M. Louis LEVY-GARBOUA

Professeur émérite d'économie à l'université Paris 1 – Panthéon Sorbonne

Membres de jury

Mme. Béatrice Boulu-Reshef, Professeur d'économie, Université d'Orléans

M. Etienne Bressoud, Directeur adjoint, BVA Nudge Unit

M. Andrew Clark, Professeur d'économie et Directeur des Affaires Internationales et Européennes, PSE

M. François Gardes, Professeur émérite, Université Paris 1 – Panthéon Sorbonne

M. Jean-Robert Tyran, Professeur d'économie, Directeur du Centre d'Économie Expérimentale et Vice-recteur Recherche et Affaires Internationales, Université de Vienne

Table of contents

| | |
|--|----|
| Acknowledgements | 5 |
| Résumé en Français de la thèse : trois essais sur l'utilisation du Nudge dans les entreprises..... | 6 |
| Introduction | 6 |
| Chapitre 1 – L'approche Nudge : un regard nouveau sur de problèmes récurrents..... | 6 |
| Chapitre 2 : Le Nudge et la participation à des enquêtes sur internet | 8 |
| Chapitre 3 : Le Nudge et la satisfaction au travail | 12 |
| Chapitre 4 : Le Nudge et la motivation au travail | 15 |
| Conclusion..... | 19 |
| Introduction | 21 |
| Chapter 1 - The Nudge approach: a fresh look at old issues | 23 |
| Introduction | 23 |
| Part 1: Old issues and limited solutions..... | 23 |
| The issues | 23 |
| The limits of monetary incentives | 24 |
| Part 2: Nudge and behavioral economics..... | 25 |
| Behavioral Economics..... | 26 |
| The Nudge approach | 28 |
| Part 3: The arguments against the Nudge approach, and how we overcome them | 29 |
| Part 4: How can the Nudge approach help with the business problems we mentioned | 32 |
| Conclusion..... | 35 |
| Chapter 2 - A Nudge is worth a thousand words: increasing web survey participation | 36 |
| Introduction | 36 |
| Part 1: Literature review | 36 |
| The literature on web survey participation | 36 |
| The Nudge approach | 38 |
| Part 2 - General study design | 39 |
| The process | 39 |
| The experiments..... | 42 |
| Part 3 - Hypotheses | 43 |
| Part 4 - Experiment 1..... | 44 |
| Method..... | 44 |
| Results..... | 47 |
| Further analysis | 52 |
| Part 5 - Experiment 2..... | 54 |
| Method..... | 54 |

| | |
|---|-----|
| Results | 55 |
| Differences between experiments 1 and 2 | 59 |
| Further analysis | 60 |
| Part 6 - Discussion | 61 |
| Future research | 64 |
| Conclusion | 65 |
| Chapter 3 - Nudge and job satisfaction: the development of a new measure | 67 |
| Introduction | 67 |
| Part 1 – Theoretical background | 68 |
| Internship or professional satisfaction? | 69 |
| Assessing job satisfaction | 69 |
| Aimed at colleges and much more? | 72 |
| Part 2 – Hypotheses | 73 |
| Evaluating the method | 74 |
| Understanding the determinants of a successful internship | 74 |
| Part 3 – Study design | 75 |
| Data collection..... | 75 |
| Incentives | 77 |
| Part 4 –Results..... | 77 |
| Descriptive statistics..... | 77 |
| Evaluating the method | 80 |
| Do people who learn a lot in the first period have a different profile? | 85 |
| The role of the activities performed | 87 |
| Understanding the determinants of a successful internship | 88 |
| Predicting final satisfaction | 91 |
| Predicting the internship grade | 94 |
| Part 5 - Discussion | 97 |
| Conclusion..... | 99 |
| Chapter 4 – Nudge and motivation at work: the case of gamification | 100 |
| Introduction | 100 |
| Part 1 - Why Gamification? | 102 |
| Using gamification to Nudge performance: A behavioral economic interpretation | 103 |
| Human motivations at work: the economic view of incentives deepened by psychology?..... | 104 |
| Motivation and performance at work | 105 |
| Part 2 - An economic model of gamification in the workplace | 106 |
| Fixed payment..... | 108 |

| | |
|--|-----|
| Piece rate | 109 |
| Part 3 - Hypotheses | 111 |
| The two stages of non-pecuniary rewards: anticipation and awareness..... | 111 |
| The hypotheses | 112 |
| Part 4 - Experimental design | 113 |
| Procedure and design of the study..... | 113 |
| The sessions and participants..... | 117 |
| Part 5 - Results | 117 |
| Effect of gamification on performance..... | 119 |
| Effect of gamification on intrinsic motivation | 124 |
| Effect of gamification on break time | 126 |
| The interaction between gamification and ability..... | 127 |
| The effects of gamification under a piece rate | 129 |
| Is gamification cost-effective?..... | 130 |
| Part 6 - Discussion | 130 |
| Limitations and further research | 133 |
| Applications..... | 133 |
| Conclusion..... | 134 |
| Conclusion..... | 136 |
| Appendix | 137 |
| References | 144 |

Acknowledgements

This thesis was not just my work. More than anything, it was a team effort. It would not be what it is today without the contribution of many other people.

I would like to start by thanking the BVA group for funding this thesis via a CIFRE contract. I would also like to thank Etienne Bressoud, PhD and Deputy CEO at the BVA Nudge Unit, who was my tutor in the company. The very generous financial contribution of BVA, as well as Etienne's continuous support and advices, are what made this thesis possible. I also extend my gratitude to all of my former colleagues of the BVA Nudge Unit, notably my friends Alice Debost, Pauline Le Golvan and Beltrande Bakoula, who contributed in making this whole experience much more pleasant.

I would also like to thank Béatrice Boulu-Reshef for all the advices she gave me. My thanks also go to my friend and fellow PhD student, Feryel Bentayeb. I worked with her extensively on the chapters 3 and 4 of this thesis, and both of these experiments are the product of our collaboration and of a great working relationship. I would also like to thank Maxim Frolov for his work in the programming of the experiment in chapter 4, and Loïc Sorel for all the administrative support and help he gave me.

On a more personal note, I would like to thank my parents for all the support they gave me. My thanks also go to Salomé Daniel, my girlfriend, who gave me the emotional help I needed during the toughest moments, and whose expertise in psychology also helped me in my work. She also provided me with motivational material that was crafted expertly to ensure that I did the best I could at all times.

Finally, I would like to express my most sincere gratitude to my thesis director, Louis Levy-Garboua. He spent a considerable amount of time guiding me, helping me, offering advices and new ideas, as well as support when needed. He went way beyond what was expected of him as a thesis director, and I could not have dreamt of a better tutor.

I thank you all again for everything you did, and I hope that the final result will live up to your expectations.

Résumé en Français de la thèse : trois essais sur l'utilisation du Nudge dans les entreprises

Introduction

Quand Richard Thaler a reçu le Prix Nobel en 2017, cela a envoyé un message fort : l'approche Nudge est là pour durer. Ce prix est la reconnaissance ultime pour les nombreux succès du Nudge depuis la publication, par Thaler et Sunstein du livre *Nudge* en 2008. La raison principale pour laquelle cette approche est aussi populaire est sans doute qu'elle est arrivée à un moment où les entreprises et les gouvernements en avaient grandement besoin.

Il s'agit en effet d'une approche versatile, peu coûteuse, et potentiellement très efficace. Dans cette thèse, nous allons d'abord présenter dans un premier chapitre les problèmes auxquels nous nous proposons de répondre, puis nous allons expliquer en détails en quoi consiste le Nudge. Ensuite, dans chacun des trois chapitres suivants, nous allons présenter un problème précis qui préoccupe les entreprises, puis nous allons montrer comment le Nudge peut être utilisé pour apporter des solutions.

Chapitre 1 – L'approche Nudge : un regard nouveau sur de problèmes récurrents

La dernière décennie a vu l'avènement d'une nouvelle approche de l'économie : l'approche Nudge. Le terme « Nudge » a été d'abord utilisé par Thaler et Sunstein dans le livre qu'ils ont publié en 2008, intitulé : « *Nudge : Améliorer les décisions concernant la santé, la richesse et le bonheur* ». Basé sur les dernières avancées en Economie Comportementale, ce livre propose une alternative aux solutions économiques traditionnelles, qui reposent généralement sur l'usage d'incitations monétaires. Au lieu de cela, il met en avant le rôle du contexte, et propose d'altérer l'environnement qui entoure une décision afin d'influencer celle-ci. En d'autres termes, alors que les différentes options restent les mêmes d'un point de vue strictement économique, elles sont présentées d'une manière différente.

Cette approche a déjà été testée avec succès dans de nombreuses situations différentes, mais nous sommes convaincus que les entreprises continuent d'être confrontées à des problèmes qui pourraient être résolus à l'aide de cette méthode alternative. Dans ce chapitre, nous présenterons

les problèmes que nous nous proposons d'adresser, puis nous expliquerons pourquoi nous pensons que l'approche Nudge pourra permettre de les résoudre.

La science économique est connue pour adresser de nombreux problèmes qui concernent tout particulièrement les entreprises, comme par exemple des questions purement économiques (le niveau de taxation optimal) ou bien des questions un peu plus psychologiques (comme la satisfaction au travail). Les problèmes purement économiques peuvent effectivement être traités via le prisme de l'approche classique, mais celle-ci pâtit dès lors qu'on tente de l'appliquer à des questions psychologiques. C'est à ces questions-là que nous allons tenter d'apporter des réponses novatrices dans cette thèse.

Le premier problème est celui de la participation à des enquêtes sur internet. Les entreprises ont besoin de pouvoir recueillir des informations sur leurs consommateurs, et certaines d'entre elles vendent même ces services à leurs clients. Un autre problème est celui de l'insatisfaction au travail. Les employés heureux sont 20% plus productifs que les autres, ce qui fait du bonheur une question économique cruciale. De même, augmenter la motivation des employés pour les tâches qu'ils seront amenés à faire est particulièrement important. Puisque ces problèmes ne sont pas récents, les économistes s'y sont déjà penchés, mais leur solution de prédilection, l'introduction d'incitations monétaires, ne fonctionne pas toujours. Ces incitations sont coûteuses, peu fiables, et peuvent même avoir des effets adverses : elles font perdre goût à la tâche (car l'individu se concentre alors sur la récompense monétaire, et non plus sur les autres formes de satisfaction que la tâche pourrait lui procurer) et elles encouragent les comportements contraires à l'éthique.

L'économie faisait traditionnellement l'hypothèse d'agents parfaitement rationnels, avec des préférences stables et complètes, qui ne s'intéressent qu'à leur intérêt propre et qui maximisent leur utilité sous les contraintes qui leur sont imposées. Cette vision des choses a perduré jusqu'aux alentours des années 50, où de plus en plus d'économistes ont cherchés à infirmer les hypothèses fondamentales sur lesquelles cette approche reposait. Puis, en 1979, Kahneman et Tversky publient un article de recherche intitulé *La théorie des perspectives*, qui démonte point par point les principes de l'utilité espérée. Ceci marque un tournant, et le début de l'Economie Comportementale en tant que courant de pensée formel, marqué par l'utilisation de la psychologie pour enrichir les hypothèses de la science économique.

C'est quelques décennies plus tard, en 2008, que Thaler et Sunstein publient leur livre, *Nudge*. Ils définissent le Nudge comme étant « un aspect de l'architecture du choix qui modifie le comportement des gens d'une manière prévisible sans leur interdire aucune option ou modifier de manière significative leurs incitations économiques. » L'architecture de choix est le contexte qui

entoure une décision donnée, et les auteurs cherchent à montrer que ce contexte, souvent négligé, a pourtant une influence cruciale. Un des exemples les plus connus est celui de la mouche dans l'urinoir : dans l'aéroport de Schiphol, à Amsterdam, ajouter une image de mouche dans les urinoirs a permis une augmentation de la propreté des toilettes de 80%. Un autre exemple est l'utilisation d'options par défaut pour faire passer le taux de participation à un plan d'épargne retraite de 20% à 90%. Cette approche est basée sur les travaux en Economie Comportementale que nous avons mis en avant lors du paragraphe précédent.

Bien entendu, une approche aussi populaire n'a pas pu échapper à un grand nombre de critiques. Par exemple, certains considèrent que l'approche n'est pas éthique car elle consisterait à manipuler les individus, tandis que d'autres trouvent que la définition est trop ambiguë, et qu'il n'est pas toujours facile de dire si une intervention donnée relève du Nudge ou pas. Adresser les questions éthiques en profondeur pourrait faire l'objet d'une thèse entière, nous dirons simplement que dans ce travail, nous ferons bien en sorte que les Nudges soient toujours des interventions douces et entièrement facultatives, en restant fidèle à l'esprit des auteurs originaux. Ainsi, les individus ne se sentiront jamais obligés, consciemment ou non, de faire ce qu'ils ne souhaitent pas faire. Concernant la définition, nous garderons la définition originale avec une seule différence : nous considérons que n'importe quel changement d'incitation monétaire, même s'il n'est pas jugé significatif, ne relève pas du Nudge. Ainsi, nous nous concentrerons entièrement sur des modifications de contexte, ce qui limitera les confusions possibles.

Cela étant dit, pourquoi considérons-nous que l'approche Nudge peut être meilleure que les incitations monétaires ? Tout d'abord, il existe de très nombreux articles de recherche montrant comment elle a pu être appliquée dans différents domaines avec succès. Cela nous permet d'avoir toute une bibliothèque de bonnes idées dans laquelle puiser pour élaborer de nouvelles interventions à tester. De plus, les Nudges se sont régulièrement avérés être plus rentables que les incitations monétaires, notamment car ils ont tendance à être bien moins chers que ces dernières tout en produisant des résultats au moins aussi forts. Enfin, ils sont beaucoup moins susceptibles de générer des effets indésirables, tels que des comportements contraires à l'éthique. Ces raisons nous semblent suffisantes pour justifier une étude approfondie de l'approche. Dans le reste de ce travail, nous allons donc l'appliquer à trois nouvelles situations : la participation à des enquêtes sur internet, la satisfaction au travail, et la motivation.

Chapitre 2 : Le Nudge et la participation à des enquêtes sur internet

Les enquêtes sur internet ont connu une hausse de popularité ces dernières décennies. Néanmoins, le taux de réponse à ces enquêtes reste inférieur de 11% à celui des autres (comme les enquêtes téléphoniques, par exemple). Cela pose de véritables problèmes méthodologiques, puisque le taux de réponse est considéré comme un critère déterminant pour la qualité d'une étude. C'est pourquoi les chercheurs ont réfléchi à de nombreux moyens d'augmenter les taux de réponse. Certains ont proposé de réimaginer la manière de rédiger le corps des e-mails, d'autres d'utiliser des rappels et des incitations monétaires, d'autres encore de personnaliser chaque message selon le destinataire (en incluant le prénom dans le titre du mail, par exemple). Ce que nous nous proposons de faire dans ce chapitre, c'est d'ajouter une autre solution potentielle à cela : le Nudge. Une autre originalité de notre travail sera de passer par un canal particulier qui n'a pas encore été étudié formellement : l'invitation d'individus qui viennent de participer à une enquête téléphonique.

Pour réaliser cette étude, nous sommes partis d'un processus déjà mis en place par un grand institut de sondages Français. Toutes les semaines, cet institut réalise des sondages téléphoniques sur un échantillon d'environ un millier de personnes représentatives de la population française. À l'issue de chaque appel, si le répondant a accepté de répondre au sondage puis est allé jusqu'au bout de celui-ci, l'enquêteur va lui proposer de répondre également à des enquêtes sur internet. Si le répondant accepte, l'enquêteur va alors noter son adresse e-mail dans une base de données. Quelques jours plus tard, le répondant recevra par e-mail une invitation à l'enquête en ligne évoquée précédemment.

Les enquêteurs suivent un script spécifique tout au long de l'enquête. Ce script inclut la manière dont ils posent la question. Ainsi, la formulation qui était utilisée avant que nous n'entamions cette recherche était ceci :

« F0a Enfin, seriez-vous d'accord pour être contacté par mail pour répondre à des études ? »

« F0b Enfin, seriez-vous d'accord pour être de nouveau contacté par téléphone pour un prochain sondage ? »

L'enquêteur commençait par demander au répondant s'il souhaitait être recontacté par mail puis, après avoir pris note de sa réponse (et de son e-mail en cas de réponse positive), il lui demandait alors s'il souhaitait être recontacté par téléphone.

Ceci présentait une opportunité de Nudge évidente : peut-être qu'en modifiant la manière dont cette demande était formulée, il serait possible d'obtenir plus de participants ? C'est pourquoi, dans cette expérience, nous avons proposé des scripts alternatifs. Les enquêteurs ont donc lu aux répondants une formulation aléatoire, sélectionnée parmi l'ensemble des traitements y compris le contrôle (qui

était la formulation utilisée auparavant, que nous avons présenté ci-dessus). Nous avons alors noté deux variables. D'abord, le taux d'acceptation : la proportion d'individus acceptant de donner leur adresse e-mail. Ce taux avoisinait les 20% en moyenne. Puis, le taux de participation : la proportion d'individus qui, après avoir donné leur adresse e-mail, ont effectivement participé à l'enquête sur internet. Ce taux avoisinait les 3%.

Nous avons donc testé un total de 10 formulations alternatives élaborées à partir de nombreuses recherches en économie comportementale. Cette expérience a été menée durant deux périodes : une première fois pendant l'été 2015, puis une seconde fois durant l'hiver 2015 et le début d'année 2016. Trois traitements sont identiques d'une période à l'autre, dont le traitement de contrôle. Ceci nous a permis de vérifier si un même traitement pouvait avoir un effet différent selon le moment où il était implémenté. En plus de ces trois traitements communs, nous avons testé quatre traitements uniques durant chacune de ces deux périodes.

Voici quelques exemples de formulations qui ont été essayées :

« F1 Merci d'avoir participé à cette enquête. Nous vous proposons de rejoindre les nombreux Français qui ont souhaité participer à des enquêtes par e-mail. Pour ce faire, merci d'épeler votre adresse mail. »

« F2 Merci d'avoir participé à cette enquête. Nous vous proposons de participer à de nouvelles enquêtes par mail. Pour ce faire, merci d'épeler votre adresse mail. »

« F3 Avez-vous trouvé que cette étude était :

1. Très intéressante.
2. Plutôt intéressante.
3. Intéressante.
4. Moyennement intéressante.
5. Peu intéressante.

[Si « *moyennement intéressante* » ou *plus*]

Dans ce cas, je vous propose de participer à de nouvelles études par mail. Pour ce faire, merci d'épeler votre adresse mail. »

La plupart des formulations que nous avons testées, y compris celles-ci-dessus, ont permis une augmentation significative du taux d'acceptation, qui est notamment passé de 24% avec le contrôle (Q0) à 30% avec la formulation Q2. Néanmoins, la plupart ont échouées à augmenter le taux de participation. Si le Nudge avait bien permis, dans ces cas, de collecter davantage d'adresses e-mail, cela n'a pas permis de donner davantage envie de participer le moment venu. Si l'expérience n'avait été menée qu'en été, alors nous en serions resté là, mais la session menée en hiver nous a permis d'apporter un éclairage intéressant.

Durant cette période, trois formulations ont eu un effet significatif sur le taux de participation. Ces formulations et les chiffres correspondants sont montrés dans le tableau ci-dessous

Tableau 1 – Effet de trois Nudges sur le taux d'acceptation et le taux de participation

| Traitement | Taux d'acceptation (%) | Taux de participation (%) |
|--|------------------------|---------------------------|
| Contrôle | 19.1 | 2.5 |
| F4 « Merci d'avoir participé à cette enquête. Si vous souhaitez continuer à donner votre avis, seriez-vous d'accord pour participer de nouveau par mail pour répondre à des études ou par téléphone pour un prochain sondage ? » | 23.1 | 4.7* |
| F5 « Merci d'avoir participé à cette enquête. Enfin, seriez-vous d'accord pour participer de nouveau par mail pour répondre à des études ou par téléphone pour un prochain sondage ? » | 24.5* | 4.5* |
| F6 « Enfin, si vous souhaitez continuer à donner votre avis, seriez-vous d'accord pour participer de nouveau par mail pour répondre à des études ou par téléphone pour un prochain sondage ? » | 27.1* | 4* |

* Significativement différent du groupe de contrôle (p -value inférieure à 10% d'après un test de Student)

Finalement, ce rend ces formulations différentes des autres est leur simplicité : plutôt que d'employer directement des leviers maintenant connus, comme les normes sociales (en mettant en avant que de nombreux Français participent déjà, par exemple¹), elles se contentent de rendre plus polie et amicale la formulation de contrôle, qui était très neutre. Mais ce n'est pas tout ce qu'il faut retenir ici.

En effet, la formulation F4 a une particularité par rapport aux deux autres : elle avait déjà été testée en été, où elle avait permis une augmentation du taux d'acceptation mais n'avait eu aucun effet sur le taux de participation. Quant aux formulations F5 et F6, il ne s'agit que des deux versions réduites de la formulation F4. Ce phénomène surprenant à premier abord illustre une vérité essentielle de l'approche Nudge : le contexte a une importance cruciale. Ici, la seule différence entre le succès et l'échec d'une formulation était la période de l'année durant laquelle elle a été testée. Pourquoi une telle différence ? Car les individus sont dans un état d'esprit différent d'une période à l'autre. En

¹ Quelqu'un pourrait penser qu'un taux de participation de 2.5% signifie que peu de Français participent, et que cette formulation est donc mensongère, mais il faut garder à l'esprit que les sondages sont menés sur 1000 personnes par semaine. Ainsi, au bout de 12 semaines d'enquêtes, c'est déjà 300 personnes qui participent, ce qui n'est pas négligeable.

effet, l'été a tendance à être une période plus agréable pour de nombreuses personnes, qui sont en vacances et qui, au moment de répondre à l'enquêteur, surestiment grandement leur capacité (et motivation) à participer plus tard, quand les vacances seront terminées et que l'e-mail arrivera. En hiver, au contraire, beaucoup d'individus sont très occupés et ont moins de temps à consacrer à des enquêtes. Dans notre étude, cela s'est traduit par un taux d'acceptation et de participation moyen significativement plus bas en hiver qu'en été. En revanche, cela implique qu'un traitement qui les remercie pour l'effort qu'ils ont fait, et qui se contente de demander humblement à l'individu s'il souhaite continuer à participer, sera d'autant plus apprécié. Et cette différence est assez pour que le taux de participation augmente significativement.

Après avoir montré comment appliquer le Nudge pour augmenter le taux de participation à des enquêtes sur internet, et avoir ce faisant découvert des phénomènes surprenants, nous allons maintenant voir comment l'approche peut être utilisée pour améliorer la manière dont la satisfaction au travail est mesurée.

Chapitre 3 : Le Nudge et la satisfaction au travail

La littérature scientifique regorge d'articles prouvant l'importance de la satisfaction au travail. En plus du fait que les employés heureux au travail sont 20% plus productifs, des employés satisfaits pourront devenir des ambassadeurs de l'entreprise, ce qui aidera ses efforts de recrutement et augmentera son taux de rétention. Cela fait de la satisfaction au travail un enjeu économique majeur. Mais avant de se demander comment l'améliorer, il faut déjà savoir comment la mesurer efficacement.

La méthode la plus populaire consiste à administrer un questionnaire aux employés, généralement une fois par an, en utilisant des questionnaires types dont la fiabilité a été prouvée par des études académiques. Parmi les plus populaires, il y a le *Minnesota Satisfaction Questionnaire* et le *Job Descriptive Index*, qui visent à mesurer les différents aspects de la satisfaction au travail en utilisant 100 et 72 questions respectivement. Les résultats peuvent s'avérer très riches, mais ces méthodes ne sont pas dénuées d'inconvénients.

Le principal problème est le fait que ces questionnaires ne mesurent pas tant ce que la personne ressent, mais ce dont elle se souvient. En effet, il est demandé à l'individu de se remémorer l'année qu'il vient de passer, et de se rappeler de points précis. Kahneman et ses collègues ont publié un célèbre papier intitulé « *Back to Bentham* » à ce sujet : ils expliquent que la satisfaction au travail est, en réalité, une évaluation de l'utilité que l'individu ressent à un moment donné. Or, ces

questionnaires mesurent une utilité « remémorée », puisqu'on demande un effort de mémoire à l'individu. Le résultat est donc différent ce que l'on souhaite mesurer au départ. En plus de cela, l'utilité remémorée souffre d'un certain nombre de biais méthodologiques. Le plus connu d'entre eux est le biais de « peak-end » : lorsqu'un individu tente d'évaluer sa satisfaction par rapport à une expérience, il se souvient essentiellement du moment le plus intense (le pic), et du dernier moment (la fin), ce qui le pousse à sous-estimer grandement l'importance du reste de l'expérience.

C'est pourquoi, dans ce chapitre, nous nous proposons d'utiliser l'approche Nudge pour concevoir une méthode de mesure alternative, qui ne souffrirait pas de ces inconvénients. Contrairement aux autres chapitres, nous n'allons pas créer de Nudges en soi, mais plutôt utiliser notre connaissance des biais de décision pour concevoir une mesure qui les prend mieux en compte. Nous allons également nous inspirer des innovations récentes en la matière, comme par exemple l'« Experience Sampling Method », qui consiste à interroger l'individu à plusieurs reprises chaque jour à des instants aléatoires pendant une période donnée, puis à lui demander comment il se sent au moment où il est interrogé.

La méthode que nous avons testée est la suivante : des étudiants de niveau Licence à Master amenés à faire un stage en 2017 ont été invités à participer à l'étude. Nous leur avons demandés de répondre à un premier questionnaire, administré juste avant qu'ils ne démarrent leur stage, pour enregistrer des variables sociodémographiques, évaluer leur personnalité et noter ce qu'ils attendent de leur stage. Ensuite, ils devaient répondre à un questionnaire supplémentaire qui leur a été envoyé une fois par mois à une date et heure aléatoire (selon un planning prédéterminé généré aléatoirement et qui leur est inconnu), afin qu'ils nous disent quelles activités ils ont effectué durant les deux jours de travail précédant la réception du questionnaire, puis à quel point ils en sont satisfaits et à quel point ils considèrent avoir appris en les faisant.

Une fois le stage terminé, ils ont répondu à un questionnaire final dans lequel ils disaient dans quelle mesure le stage a répondu à leurs attentes, puis ils nous ont transmis leur note de stage. L'objectif de cette étude était d'aller plus loin que de simples mesures générales de satisfaction. En interrogeant les étudiants sur les activités qu'ils font, puis en les interrogeant de manière régulière, nous pouvons alors comprendre plus précisément ce qui, dans leur stage, leur plaît et les dérange. Nous pouvons aussi savoir quelles activités sont les plus enrichissantes pour eux, à quels moments et pourquoi certains perdent l'intérêt qu'ils ont pour leur stage, par exemple. De plus, le fait d'interroger plusieurs fois durant le stage, plutôt qu'à la fin, et en mentionnant les deux derniers jours, une période qui reste très courte, il est possible d'éviter les écueils des mesures classiques.

Utiliser à la fois cette mesure novatrice et la mesure classique de satisfaction (questionnaire plus détaillé en fin de période), cela nous permet de comparer les deux en termes d'enseignement.

Cette étude nous a permis de constater l'existence d'un certain nombre de phénomènes. Tout d'abord, la satisfaction est très corrélée au sentiment d'apprentissage. Ce résultat est assez évident dans le contexte d'un stage, où les individus ont pour objectif principal d'apprendre. Une conséquence de cela est que les activités qui leur permettaient d'apprendre le plus étaient aussi celles dont ils étaient le plus satisfaits. Néanmoins, le seul résultat que nous avons pu tirer d'une analyse par activités était que les tâches administratives étaient moins gratifiantes que les autres.

Nous avons ensuite effectué une série de tests de corrélation afin d'observer les liens entre les différentes variables. Les résultats sont consignés dans le tableau ci-dessous :

Tableau 2 – Tests de corrélation entre les différentes mesures de satisfaction au travail et la note de stage

| | Satisfaction début | Satisfaction milieu | Satisfaction fin | Satisfaction finale |
|---------------------|--------------------|---------------------|------------------|---------------------|
| Satisfaction début | | | | |
| Satisfaction milieu | 0,45** | | | |
| Satisfaction fin | 0,36* | 0,60*** | | |
| Satisfaction finale | 0,60*** | 0,27 | 0,37* | |
| Note de stage | 0,07 | 0,04 | 0,07 | 0,42* |

$p < .001$ '***', $p < .01$ '**', $p < .05$ '*'

La majorité des sujets de l'expérience ont effectué un stage de six mois, mais d'autres ont eu un stage un peu plus court. Pour améliorer la qualité de l'analyse, nous avons donc opéré à une transformation des données afin d'obtenir, pour chaque individu, trois mesures de satisfaction : la satisfaction au début du stage, la satisfaction au milieu, et celle à la fin.

Plusieurs résultats émergent : tout d'abord, les résultats mensuels sont très corrélés les uns aux autres, ce qui indique que la satisfaction d'un mois donné permet de prédire dans une certaine mesure la satisfaction des autres mois. Par ailleurs, on constate que la dernière mesure de satisfaction est corrélée avec la satisfaction finale, ce qui est une confirmation de l'effet de « peak-end ». La satisfaction du « milieu » de l'expérience n'est pas corrélée avec la satisfaction finale. En revanche, la corrélation entre la satisfaction finale et la satisfaction du début du stage est très forte. Ce résultat illustre bien l'importance cruciale du début du stage. Il est primordial que le stage se passe bien dès le départ, car sinon cela aura un impact fort sur ce dont l'individu se souviendra à la fin du stage. Cela implique aussi qu'il est possible de prédire la satisfaction finale dès le premier mois.

C'est sans doute le résultat le plus fort de cette recherche. Dans ce travail, le résultat des enquêtes de satisfaction n'était pas communiqué à l'entreprise, mais si c'était le cas ? Cela signifie qu'avec un questionnaire court et simple administré très tôt, il est possible d'avoir une mesure fiable de la

satisfaction dès le premier mois. Ceci qui permet alors au manager ou aux ressources humaines d'agir immédiatement pour détecter les problèmes et améliorer la satisfaction au travail d'un employé avant qu'il ne soit trop tard. Il n'est plus nécessaire d'attendre toute une année avant de réaliser que quelque chose ne va pas. Bien sûr, nous avons étudié le cas des stagiaires, mais celui d'employés en début de carrière nouvellement recruté sera très similaire : ils se trouveront dans un environnement de travail nouveau, avec comme objectif principal d'apprendre. Ainsi, nous recommandons vivement aux entreprises menant des campagnes de recrutement de considérer l'utilisation d'une mesure simple comme celle-ci, en complément des méthodes de satisfaction traditionnellement employées.

Maintenant que nous avons vu comment l'approche Nudge peut être appliquée pour repenser la manière de mesurer la satisfaction au travail, nous allons voir comment elle peut être utilisée pour améliorer la motivation au travail.

Chapitre 4 : Le Nudge et la motivation au travail

La satisfaction au travail est un aspect déterminant de la productivité. La motivation en est un autre. Pour mieux comprendre le problème, il convient d'abord de distinguer deux formes de motivation : la motivation intrinsèque et la motivation extrinsèque. La première désigne la motivation qui vient de l'intérieur d'un individu. Par exemple, quand l'individu effectue une tâche car il la trouve intéressante ou parce qu'il apprécie le défi qu'elle représente. Dans ce cas, l'individu s'intéresse à la tâche en elle-même. La seconde désigne la motivation qui vient de l'extérieur de l'individu : celui-ci fera la tâche parce qu'il s'attend à recevoir une récompense externe, qui pourra être de l'argent, mais aussi un meilleur statut social. Dans ce cas, l'individu ne s'intéresse pas à la tâche, mais plutôt à ce que celle-ci lui permettra d'obtenir. Typiquement, les incitations monétaires jouent sur la motivation extrinsèque, tandis que les efforts pour rendre le travail plus épanouissant jouent sur la motivation intrinsèque.

Dans ce chapitre, nous nous pencherons plus particulièrement sur le cas des employés consacrant une grande partie de leur journée à des tâches routinières et répétitives. Naturellement, il peut être très difficile de trouver de la motivation intrinsèque pour ce genre de tâches, c'est pourquoi les entreprises sont à la recherche de solutions pour les rendre plus intéressantes. Comme nous l'avons expliqué dans le chapitre 1, les incitations monétaires ne sont pas toujours une réponse fiable à ce genre de problème, notamment car elles sont chères, peu fiables, et peuvent même s'avérer contreproductives.

La « gamification » se veut être une réponse moderne à ces difficultés. La gamification est définie comme étant l'utilisation d'éléments de « game design » (conception de jeux) dans des contextes non ludiques. En d'autres termes, il s'agit de voir ce qui rend les jeux (et plus particulièrement les jeux vidéo) amusants, puis de transposer ces éléments dans le monde professionnel afin de rendre celui-ci plus stimulant. La gamification est très souvent une application de l'approche Nudge : en effet, le concepteur va typiquement changer la manière de présenter une tâche de manière à ce qu'elle soit perçue comme plus intéressante, sans que cela ne change les incitations monétaires de quelle que manière que ce soit. L'hypothèse fondamentale est que si la tâche est rendue plus amusante, alors ceci permettra d'améliorer la motivation intrinsèque de l'individu pour cette tâche. La réalité est un peu plus complexe. En effet, il y a de nombreuses manières différentes de « gamifier » une tâche, et certaines jouent davantage sur la motivation extrinsèque que sur la motivation intrinsèque. Une méthode populaire déjà employée avant même l'apparition du terme de « gamification » consiste à donner des récompenses symboliques aux employés les plus performants. On peut penser par exemple au concept d'employé du mois.

Une des faiblesses de la littérature sur la gamification est d'ailleurs son manque de distinction entre les implémentations qui consistent uniquement en l'ajout de badges et d'un tableau des scores, et celles qui tentent de modifier la présentation de la tâche en elle-même. Les approches basées essentiellement sur l'ajout de badges ont d'ailleurs été labellisées de « pointification », pour montrer que leurs concepteurs ne s'intéressent pas vraiment aux éléments fondamentaux qui rendent les jeux amusants, mais en prennent simplement des constituants superficiels. Pour bien marquer cette distinction, nous parlerons à présent de « récompenses non-pécuniaires » pour désigner ces interventions visant à améliorer la motivation sans recourir à des incitations monétaires (qui sont donc des récompenses pécuniaires).

Ainsi, dans ce chapitre, nous allons tester ces deux manières d'implémenter des récompenses non-pécuniaires, à savoir la pointification et la gamification, et nous allons comparer ceci à l'utilisation d'incitations monétaires. Nous allons ensuite mesurer la motivation intrinsèque des individus, ainsi que leur productivité, ce qui nous permettra de savoir quelle méthode est la plus efficace par rapport à ces deux éléments.

Pour ce chapitre, nous avons décidé de mener une expérience en laboratoire. En effet, la motivation intrinsèque est un sujet très complexe qui peut être influencé par une myriade de facteurs connexes, comme la bonne entente avec les collègues ou les vocations d'un individu. Faire un test en laboratoire, plutôt que sur le terrain, permet donc de suffisamment simplifier le problème pour que nous puissions tirer des conclusions claires de nos résultats.

Les sujets que nous avons invités ont dus effectuer une tâche simple et répétitive pendant un total de trente minutes. La tâche que nous avons choisi de leur demander consistait à reconstituer des mots à partir d'un ensemble de lettres qui s'affichait à l'écran. Par exemple, l'écran afficherait « riavo », et le sujet pourrait reconstituer le mot « avoir ». Une fois que c'était fait, le sujet pouvait passer au mot suivant. Et s'il n'y arrivait pas, il passait automatiquement au suivant après 60 secondes. Ceci était répété pendant trente minutes. Tous les 10 ensembles de lettre, ce que nous appelons un « écran de transition » était affiché : cet écran indiquait le nombre de mots que le sujet avait réussi à reconstituer jusqu'à présent, ainsi que d'autres informations dépendant du traitement.

Nous avons choisi cette tâche en particulier car nous voulions qu'elle soit représentative d'un métier de service (en faisant fi de l'aspect social de ce genre d'emploi le temps de notre expérience) : des problèmes arrivent régulièrement, ils sont de difficulté variable mais toujours relativement simples, et le sujet doit en régler autant que possible dans le temps qui lui est imparti. Contrairement à la plupart des métiers consistant à effectuer des tâches simples, celui-ci est plus difficile à automatiser, ce qui évitera donc aux résultats de notre expérience de devenir rapidement obsolètes.

Une fois la tâche terminée, nous avons demandé aux individus de répondre à un questionnaire mesurant leur motivation intrinsèque pour ce qu'ils viennent de faire. Nous avons utilisé pour cela le questionnaire intitulé Intrinsic Motivation Inventory.

Concernant les traitements, nous avons utilisé un design 2 * 3 : deux types de récompenses pécuniaires (salaire fixe ou taux à la pièce), et trois types de récompenses non-pécuniaires (aucune, gamification ou « ville », pointification ou « badges »). Dans le traitement « salaire fixe », les sujets ont touchés 8 € à la fin de l'expérience peu importe leur performance, tandis que dans le traitement « taux à la pièce », ils étaient rémunérés de 10 centimes par mot trouvé. Dans l'écran de transition, les sujets étaient informés du montant actuel de leur rémunération.

Dans le traitement de pointification, aussi appelé « badges », le sujet pouvait gagner des médailles virtuelles purement symboliques au fil de sa progression. Tous les 10 mots trouvés, il gagnait une médaille. La quantité totale de médailles que le sujet avait obtenu était affichée sur l'écran de transition. Dans le traitement de gamification, aussi appelé « ville », tous les 10 mots trouvés, le sujet contribuait à la construction d'une ville virtuelle. Au début de l'expérience, l'écran de transition affichait un terrain vide, et plus le sujet reconstituait de mots, plus la ville se remplissait. Avec ce traitement, nous espérions que ceci donnerait l'impression que la tâche a du sens, puisqu'elle contribue à construire quelque chose. Ainsi, nous avons fait l'hypothèse que ceci améliorerait la motivation intrinsèque du sujet.

Les deux images ci-dessous représentent ce que voyaient les sujets au début et à la fin de l'expérience (à supposer qu'ils aient réussi à atteindre le niveau maximal).

Figure 1 – L'écran de transition du traitement « badges » après la première réussite (à gauche), et après la dernière (à droite)



Figure 2 – L'écran de transition du traitement « ville » après la première réussite (à gauche), et après la dernière (à droite)



Lors de l'analyse des données, nous avons déjà cherché à savoir si la gamification et la pointification avaient effectivement pu améliorer la productivité des sujets. Le tableau ci-dessous indique la performance moyenne des sujets dans chaque traitement, ainsi que la rémunération moyenne.

Tableau 3 – Performance et coût moyen dans chaque traitement

| Traitement | Performance moyenne (mots trouvés) | Coût moyen par sujet (€) |
|----------------------------|------------------------------------|--------------------------|
| Salaire fixe + contrôle | 77 | 8 |
| Salaire fixe + badges | 86 | 8 |
| Salaire fixe + ville | 91* | 8 |
| Taux à la pièce + contrôle | 93* | 9.3* |
| Taux à la pièce + badges | 84 | 8.4 |
| Taux à la pièce + ville | 87 | 8.7 |

*: Significativement supérieur au traitement «salaire fixe + contrôle » d'après un test de Student (p -value < 0.10)

Plusieurs résultats émergent de ce tableau. Tout d'abord, on peut voir que, par rapport à un traitement de contrôle, le traitement « ville » a un effet sur la productivité équivalent à l'ajout d'incitations monétaires. La gamification atteint la même hausse de performance que les incitations. Mais, naturellement, la gamification ne requiert pas d'augmenter le coût par sujet pour arriver à ce résultat. Ce tableau suggère aussi que jouer sur la motivation intrinsèque fonctionne mieux que de jouer sur la motivation extrinsèque (avec les badges), car la pointification n'a pas fonctionné.

Un autre résultat intéressant est le fait que, si la gamification et les incitations monétaires parviennent à augmenter significativement la performance quand elles sont utilisées en même temps, l'effet disparaît quand elles sont combinées. Ceci montre bien qu'on ne peut pas simplement ajouter la gamification par-dessus des incitations monétaires : il faut choisir l'un ou l'autre.

Ensuite, nous avons testé l'impact de la gamification sur la motivation intrinsèque. Étonnamment, aucun des traitements n'a eu d'effet significatif sur le plaisir que les sujets ont pris à faire la tâche. En revanche, le traitement « salaire fixe + ville » (et seulement celui-ci) a conduit à une augmentation significative du sentiment de compétence par rapport au traitement de contrôle. Une régression linéaire a ensuite révélé que c'est cet effet qui était à l'origine de la différence de productivité entre « salaire fixe + ville » et le traitement de contrôle. En d'autres termes, pouvoir voir leur ville progresser petit à petit donnait aux sujets le sentiment qu'ils étaient plutôt compétents, puisqu'ils arrivaient à accomplir rapidement quelque chose de significatif (à savoir ajouter de nombreux bâtiments de mieux en mieux). Il est intéressant de constater que nous ne retrouvons pas le même effet dans le traitement « taux à la pièce + ville ». La raison la plus probable est que le fait que le gain monétaire par mot reste faible, et donc modère l'enthousiasme du sujet. Une autre possibilité est tout simplement qu'en la présence d'incitations monétaires, alors le sujet va toujours se concentrer sur ces dernières, et donc ne prêtera plus attention à la gamification, qui ne sera plus perçue que comme une distraction.

Au final, comme la gamification est tout aussi efficace mais moins coûteuse et moins risquée, nous recommandons aux entreprises de d'abord tester cette méthode pour améliorer la motivation et la productivité de leurs employés. Si cela ne fonctionne pas, il sera toujours possible de passer aux incitations monétaires. En revanche, enlever ces dernières pour les remplacer par de la gamification sera très probablement mal vécu. Et dans tous les cas, mieux vaut choisir une de ces méthodes et s'y tenir plus que d'essayer d'implémenter les deux en même temps.

Conclusion

Ce que nous avons présenté dans cette thèse n'était qu'une petite sélection d'exemples de Nudges. Néanmoins, nous espérons qu'ils ont été riches d'enseignement et que nous sommes parvenus à montrer que cette approche pouvait être utilisée dans des cas très différents les uns des autres. En effet, le but final de ce travail n'est pas juste de montrer comment améliorer le taux de participation, la manière de mesurer la satisfaction, et la motivation au travail. L'intérêt de cette thèse est avant tout de proposer une manière de penser les choses, une méthode pour résoudre les problèmes en les attaquant sous un angle différent de ce qui se faisait jusqu'à présent. Peu importe le problème, la méthode est toujours plus ou moins la même : acquérez une bonne compréhension du problème (via une revue de la littérature, ou simplement en collectant des témoignages d'individus sur le terrain), cherchez de l'inspiration du côté des Nudges qui ont déjà été implémentés par le passé pour essayer de trouver des manières créatives de résoudre des problèmes similaires au vôtre, et ensuite voyez comment adapter ces solutions à votre cas particulier. Nudger, c'est avant tout prêter une attention extrême aux détails et au contexte. Après tout, dans le chapitre 2, la différence entre un Nudge réussi et un Nudge raté tenait en quelques mots. Une fois que vous avez trouvé les Nudges que vous souhaitez tester, il ne reste plus qu'à expérimenter, apprendre, et s'adapter. Sur ce, nous espérons que ce travail vous a donné envie de tester le Nudge vous-même, et nous vous souhaitons bon courage dans vos efforts !

Introduction

The attribution of the Nobel Prize to Richard Thaler in 2017 sent a powerful message: The Nudge approach is here to stay. This prize was the ultimate recognition of the many successes of Nudging since the publication, by Thaler and Sunstein, of the book *Nudge* in 2008. In 2002, Daniel Kahneman got the Nobel Prize for his work with Amos Tversky on Prospect Theory, published in 1979. But the popularity of Behavioral Economics and Nudge didn't wait for these Nobel Prizes to explode. The appeal of cheap interventions that result in impressive changes in behavior is obvious, and the approach is versatile enough to have a large range of applications.

Perhaps the main reason why this approach was so successful is that it was greatly needed. It arose because of the flaws of the classical approach to economics and as an alternative for situations where the old solutions didn't work. In this work, we want to illustrate the interest of Nudging to effectively influence behavior at a minimal cost, and how it can be used in different situations. We also want this work to be directly useful to decision-makers in businesses and organizations. We will deliberately limit ourselves to behavioral problems that are relevant to business, and we will provide decision-makers with the actionable insights they need to turn our results into applications.

This thesis will be divided into four chapters. In the first one, we will start by exposing some of the issues that the classical economic approach struggles with. Then, we will present a brief history of Behavioral Economics and why it arose as a response to the issues of the classical science. Then, we will see how the Nudge approach was created as an application of Behavioral Economics. We will end this chapter by explaining why we believe that this approach is particularly relevant to the problems that businesses still face today.

In the second chapter, we show our first application of the Nudge approach: increasing web survey participation. As we will show in this chapter, the participation rates are dwindling and the classic economic approaches often fail to yield significant results. We will see that Nudges managed to significantly increase web survey participation, and we will see why context is of utmost importance when using this approach. The results we obtain will prove useful for anyone wanting to maximize the participation to their web surveys, or merely wanting to collect more e-mail addresses.

In the third chapter, we will use modern behavioral science to re-think the way satisfaction is measured in companies. This will not be an application of Nudges per-se, but rather of the underlying science behind the approach. We will take what Behavioral Economics and Nudges teach us about human nature and develop a measure of satisfaction that tries to better account for it in a way that the old measures do not. We will then test this approach on interns, and draw conclusions both for

decision-makers wanting a different way of measuring satisfaction, and for educational facilities that want to provide a better experience for interns.

In the fourth and final chapter, we will directly test the effectiveness of Nudges at increasing productivity in a simple laboratory task. We will use it to make a repetitive task more fun in a process called “gamification”. We shall see how it made employees more motivated, as illustrated by an increase in both productivity and intrinsic motivation. We will extensively compare Nudges to monetary incentives, so that we may show how and when it is better than the latter. This chapter should help decision-makers find the best way to motivate their employees based on their specific context.

Now that we presented the general plan of our work, we will move on to the first chapter.

Chapter 1 - The Nudge approach: a fresh look at old issues

Introduction

The last decade has seen the rise of a new approach in economics: The Nudge approach. The term “Nudge” was coined by Thaler and Sunstein in the book they published in 2008 titled “Nudge: Improving Decisions about Health, Wealth, and Happiness”. Informed by the latest advances in Behavioral Economics, it proposes an alternative to the more traditional economic solutions, which usually revolve around incentives. Instead, it emphasizes the role of context, and proposes to tweak the environment surrounding a decision in order to influence it. In other words, the options remain the same economically speaking, but they are presented in a different way.

This approach has already been successfully tested in many scenarios, but we believe that there are still organizational problems that could greatly benefit from this alternative. In this chapter, we will present the issues we want to tackle and why we think the Nudge approach will be useful in solving them.

In the first section of this chapter, we will analyze in more detail the issues that we mentioned in the previous paragraph, as well as the shortcomings of the current solutions. Then, in the second section, we will present the Nudge approach in detail, and show how it can help steer behavior in the right direction. In the third section, we will summarize the main criticisms against the Nudge approach and how we aim to overcome them. Finally, in the fourth section, we will explain why the Nudge approach is relevant to the issues we raised in the first section.

Part 1: Old issues and limited solutions

The issues

Economic science is well-known for being able to tackle many issues that are important for business, ranging from obviously economic issues like the optimal level of taxation to more psychological ones like employee satisfaction. Many issues of interest to business firms can be studied and solved by a conventional economic approach. These are prices, taxation, wages, contracts, etc. For example, Hasset and Hubbard (2002) provide an overview of how taxation can impact business investments, while Zúñiga-Vicente *et al* (2014) examined to what extent R&D subsidies impact R&D investment in businesses (they showed that the effect can be quite sizable).

However, the conventional economic approach has been unable to deal with many other issues relevant to the management of business firms that imply psychology and bounded rationality. This is the kind of issues that will be studied in this thesis. To be more specific, we shall study three such topics: firms' acquisition of knowledge on customers' demand for their products (chapter 1) and on the satisfaction of their employees at work (chapter 2), and firms' strategies for enhancing the motivation of their employees (chapter 3).

Figuring out how to obtain information on customers' demand for their products through survey participation is an important topic for businesses. Participation to surveys of all kinds have been steadily decreasing by about 1.5% per year over the past decades (Curtin *et al*, 2005, Anseel *et al*, 2010, Schoeni *et al*, 2013), which makes it harder to evaluate the satisfaction and opinions of people. This can raise a number of problems for businesses: first, it is harder to assess what their clients or their employees think of them. Second, there are businesses whose entire model relies on selling the results of surveys to their clients. Third, it also affects the academic world, where many researches (including our own) rely on people's participation to surveys.

Another issue is job satisfaction. Indeed, happy employees are 20% more productive than unhappy ones (SGROI, 2015), and a Gallup report estimated that disengaged employees cost the U.S. \$483 billion to \$605 billion each year in lost productivity (O'Boyle & Harter, 2017). This is why businesses are always on the lookout for new ways to improve their employees' satisfaction and motivation.

Of course, since these issues are not recent, many people have already tried to come up with solutions. Economists often consider that behavioral problems such as these can simply be solved with the right monetary incentives. Give people a raise, or a performance bonus, and suddenly they'll be happier and will care more about the job. Give people money for taking surveys, and suddenly they'll want to do it. However, this method is obviously costly, but it can also prove unreliable and even backfire We will see why in the next subsection.

The limits of monetary incentives

The solutions tend to always gravitate around the same thing: incentives. It is well-known that "economics is all about incentives" (Klor *et al*, 2014). Even management has often been about finding the correct incentive scheme to motivate people to do better and be more satisfied. Of course, there is a reason why incentives are so widely used: they often work. Jenkins *et al* (1998) conducted a meta-analysis of 39 studies and concluded that monetary incentives significantly

improve performance. However, while monetary incentives are good in some cases, there are also many in which they are inefficient, or even harmful. There is now a whole body of literature showing the problems associated with monetary incentives, in particular with respect to motivation.

First, they are unreliable. Baruch and Holtom (2008) performed a meta-analysis of academic surveys, and found no correlation between the use of monetary incentives and response rate. Given that monetary incentives cost a lot of money to implement, their unreliability is a considerable problem. But this isn't the worst that can happen. Because not only can it fail to increase productivity, satisfaction or motivation, but it can also make things worse.

The most notable example of that is the crowding-out effect: this refers to the fact that adding an extrinsic reward (the incentive) for a task for which a person is intrinsically motivated can harm this intrinsic motivation. And, as a result, continuous incentives will have to be maintained, because if they ever are removed, productivity will be worse than it was before they were first introduced. This was presented very well in a meta-analysis by Deci, Koestner & Ryan (1999), who showed how it negatively impacted intrinsic motivation in a series of laboratory and field experiments. The effect was first observed by Deci (1971), who compared two groups who had to solve simple puzzles for three consecutive periods. The treatment group received a monetary incentive in the second period, while the control group did not. They observed that, during the 3rd period, intrinsic motivation of the treatment group was significantly less than that of the control group, as measured by the time that they spent engaging on the activity when there was no longer any external reward and they could leave at any time.

There is also evidence of this effect in the workplace. For example, Georgellis, Iossa and Tabvuma (2010) found that increasing extrinsic rewards in public service lowered intrinsic motivation, because it attracted people who were interested in the money rather than in public service. It can also increase the likelihood of unethical behavior. Sauer *et al* (2018) showed that when managers received a monetary incentive based on quantitative goals, they were more likely to falsify the performance reports.

In light of these issues, we believe it is time to adopt another approach to solve these problems: something that would be as effective but less costly and risky. Fortunately, such a thing already exists: it is the Nudge approach.

Part 2: Nudge and behavioral economics

Now that we have seen the limitations of the traditional monetary incentives, we will present the alternative that we propose to study in this work: The Nudge approach. In this second part, we will start by presenting a short history of Behavioral Economics and how it led to the emergence of the Nudge approach. Then, we will present the Nudge approach itself.

Behavioral Economics

The traditional way to do economics is to make the assumption of perfectly rational agents with complete and stable preferences, who only care about their own interest, and who maximize their utility under the constraints that are imposed to them. This view dominated the first half of the 20th century (although some authors, like Keynes, already included some psychological elements in their theories). While some economists truly believed this view of economic agents, most of the people defended this theory because they considered it to be easy to formalize and more relevant in practice (Mullainathan and Thaler, 2000). It was also considered that, even if people don't literally know the formulas and probabilities that rational agents are supposed to know, in practice, individuals act *as if* they knew them (Friedman and Savage, 1948). Behavioral economics arose to contradict these views.

It is in the mid 1950's that the first major criticisms of the traditional paradigms arose.² Among the early critics of the traditional paradigm, Allais (1953) mentioned the "distortion of objective probabilities and the appearance of subjective probabilities" (p.504) and said that the utility function, as traditionally described, could not account reliably for our own preferences and satisfaction. Markowitz (1952) questioned the shape of the utility function and introduced "inflection points". For instance, in gains, his function is first convex and then concave. Strotz (1955) analysed the anomalies in dynamic utility maximization, and more particularly the case of spendthrifts. Ellsberg (1961) mentioned ambiguity, and studied those situations in which agents only have imperfect knowledge about the probability associated to different outcomes. Simon (1955) questioned the excessively high computational abilities that the rational agent is assumed to have, and proposed a more realistic conception of rationality, known as "bounded rationality", to account for our cognitive limitations.

² Some authors consider that Adam Smith, with his *Theory of moral sentiments* (1759), was actually the first behavioral economist (Camerer, Loewenstein, & Rabin, 2011, Ashraf, Camerer, & Loewenstein, 2005). For instance, he says that "'we suffer more... when we fall from a better to a worse situation, than we ever enjoy when we rise from a worse to a better'" (Smith, 1759, p.311), which refers to a concept we now call "loss aversion".

Despite all of these early contributions, it is only with the publication of Prospect Theory by Kahneman and Tversky (1979) that behavioral economics really became popular. This paper presented a series of violations of expected utility (some of which analyzed earlier by Markowitz and Allais) via easily replicable experiments, and used evidence from psychology to propose a new economic theory that explains these violations. The paper promoted both the development of behavioral economics and experimental economics. The Prospect theory and the earlier papers established a guideline for researches in behavioral economics. Camerer, Loewenstein and Rabin (2011) describe it this way:

“First, identify normative assumptions or models that are ubiquitously used by economists, such as Bayesian updating, expected utility and discounted utility. Second, identify anomalies—i.e., demonstrate clear violations of the assumption or model, and painstakingly rule out alternative explanations (such as subjects’ confusion or transactions costs). And third, use the anomalies as inspiration to create alternative theories that generalize existing models. A fourth step is to construct economic models of behavior using the behavioral assumptions from the third step, derive fresh implications, and test them.” (p.6)

Behavioral economics is linked closely to experimental economics, in the way that most violations are demonstrated through experiments, both in the laboratory and in the field. Laboratory experiments are particularly relevant since they allow us to observe behavior precisely while controlling for many contextual factors. This ensures that deviations from the standard theory are due to decision-making biases and rather than to other variables. With time, this approach allowed us to question all principles of economic rationality (Thaler, 1987). In addition to the biases identified in Prospect Theory, there was also work on other variables neglected by the classical theories that individuals consider when they make decisions, such as mental accounts (Thaler, 1985), fairness (Fehr and Schmidt, 1999) and affect (Loewenstein, 2001).

The reason why this field of research is so important is that psychology is, indeed, relevant to the economic context (Rabin, 1998), even if the biases that it identified were to concern only a small number of economic agents (Akerlof and Yellen, 1985). For that reason, the findings of Behavioral Economics have started to replace the traditional neoclassical framework in many domains, for instance to analyze travel behavior (Avineri, 2012) or to design food policies (Liu *et al*, 2014).

In fact, when it comes to the application of Behavioral Economics to “real-life” problems, the findings of the field are often used indirectly, via the prism of another approach: The Nudge approach.

The Nudge approach

In 2008, Thaler and Sunstein published “Nudge: Improving decisions about health, wealth and happiness”. This book is the materialization of the philosophy they developed in an earlier paper (Sunstein and Thaler, 2003), that they call “libertarian paternalism”. Libertarian paternalism is a form of paternalism that respects freedom of choice and autonomy, and thus would be deemed acceptable by libertarians. It is an alternative to traditional paternalism, which works by limiting the freedom of individuals (via laws for instance) or using significant monetary incentives. This is not unlike asymmetric paternalism (Camerer *et al*, 2003), which consists in designing policies that help people who are acting irrationally while not harming people who act rationally. Indeed, people who are rational in the traditional sense of the term can simply ignore nudges since they preserve freedom.

In their book, they define a nudge as “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler and Sunstein, 2008, p.6). The choice architecture is the context in which the agent makes the decisions, and one of the aims of the authors is to show how the context surrounding an economic decision can have a great impact on it. The authors then show various examples of what, according to them, qualifies as a nudge. The most famous example is probably the fly in the urinal: in the Schiphol Airport, in Amsterdam, the picture of a black fly has been etched in the urinals of the men’s bathrooms. According to the authors, “it seems that men usually do not pay much attention to where they aim, which can create a bit of a mess, but if they see a target, attention and therefore accuracy are much increased” (Thaler and Sunstein, 2008, p.4). This simple nudge reduced spillage by 80%.

To show how this applies to the economic context, the authors notably show examples of saving plans for retirement. Madrian and Shea (2001, cited in Thaler and Sunstein, 2008), show that when people are not enrolled by default into a saving plan, 20% enroll after 3 months of employment, and 65% after 36 months. However, when people are enrolled by default (even if they can very easily opt-out), the enrolment of new employees increases to 90% at the beginning, and reaches 98% of people enrolled after 36 months. Since the dropout rate after one year of study only increased by about 0.5%, this nudge did not go against people’s preferences. This illustrates the power of inertia. Carroll *et al*. (2009) show that when a company adopts “forced choosing”, which means removing the default option and forcing people to choose between enrolling or not enrolling, the proportion of people who enroll increases by 25%. On the same topic, it has been found that increasing the number of different saving plans available decreases the proportion of people who

enroll, because this makes them more confused: the choice becomes harder (Iyengar, Huberman and Jiang, 2004, cited in Thaler and Sunstein, 2008).

The effective manipulations of the choice architecture are based on the findings of Behavioral Economics. The core idea is that, since Behavioral Economics have shown that we are “fallible human beings” but “predictably irrational” (Ariely, 2008), it is possible to reach more efficient outcomes using policies that are both cheaper than monetary incentives and less restrictive than laws. The job of the choice architect is thus to explore the academic literature about Behavioral Economics and Psychology, look for systematic deviations from the classical economic rationality, and then use these deviations to design cheap but effective manipulations of the *context* of decision, in order to steer the behavior of individuals toward more efficient choices (as defined by the choice architect). It is an application of Behavioral Economics in the sense that these modifications of context would not have any effect on the behavior of rational economic agents.

Nowadays, there is a lot of evidence on the many ways in which people systematically deviate from the axiomatic theory of rational choice, as well as many papers that aim to show how such deviations, or biases, can be used to improve welfare³. Many biases (default effect, anchoring, framing effect, peer effect, ...) have been identified, and Nudges explicitly take advantage of (or counter) one or several of these biases in order to orient decisions. We will give a few more examples in the final part of this chapter, but first we will present an overview of the most common criticisms against the Nudge approach and how we aim to address them in our work.

Indeed, the approach generated a lot of discussions in the academic world about the concept of nudge and the underlying ideology of libertarian paternalism. Some are concerned about the ethics of such an approach, while others criticize the fuzziness of the concept and its justification, going as far as to reject the very evidence on which it is based. Going through these arguments now is important, as it will allow us to review the evidence in favor (and against) Nudging with a more acute critical eye, and to better justify the use of the approach despite everything that could be said against it.

Part 3: The arguments against the Nudge approach, and how we overcome them

With the exploding popularity of the Nudge approach, it did not take long for a number of criticisms to arise.

³ In fact, there are a lot of papers proposing interventions that could be easily labeled as Nudges, even though the approach is not mentioned.

Many people criticized the Nudge approach for being unethical. For example, it has been argued that, while it respects freedom from external constraints (such as a strict obligation), it does not respect freedom from internal constraints such as ignorance or fear (Goodwin, 2012). Indeed, Nudges are often called “manipulative” because they supposedly take advantage of decision-making biases without the knowledge of the individual in order to change his choices and preferences. This is a simplifying view, as the variety of nudges that exist is such that only some of them could fit this description. According to Wilkinson (2013), manipulation is the combination of two things: an actor that has the intention to manipulate another person, and the perversion of the decision-making process. But is there really the intention to manipulate in the choice architect’s mind? Nudges, by definition, come with an easy opt-out option. This means that “nudges would not take the form “A intends B to do X” but rather “A intends B to do X unless B wants not to”” (Wilkinson, 2013, p.352). The choice architect could only be accused of being manipulative if he made the opt-out option very costly, so that the individual would have to let the Nudge guide him whether he wants it or not. There is not always perversion of the decision-making process either. In fact, Nudges are supposed to “make choosers better off, as judged by themselves” (Thaler and Sunstein, 2008, p.5) so, if anything, they actually improve the decision-making process by helping people stick to their preferences in spite of the biases that affect them.

Another criticism is that Nudges supposedly work best when people are not aware of them (Grüne-Yanoff, 2012). They would be manipulative in essence because they rely on the ignorance of the person being nudged to work. However, the first evidence on the matter tends to indicate that Nudges remain effective even when people are clearly told that they are going to be Nudged (Loewenstein *et al.* 2014).

This is only a short summary of the main ethical issues that are raised. It would take an entire thesis to make a full review of the ethical implications of Nudging. We simply wanted to show that this debate existed and give an overview of the main arguments, while also showing that Nudging is not as manipulative as it would seem at first glance. Our stance regarding ethics in this work is to simply stick to Nudges that will do no harm to individuals. We aim to remain as factual and objective as possible, and ethical considerations are by nature very subjective. We encourage readers who want to explore the ethical questions further to read the references we mentioned so far in this part, as they would make a good starting point.

Another issue that is more relevant to our work is the fuzziness of the definition. The definition of the Nudge has raised some problems since the publication of the book. There is no consensus around it and people often mistakenly qualify as “Nudges” interventions that do not fit

the definition. There is not even a unanimous definition of Nudging. The confusion arises from the book itself. The authors define a Nudge as “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler and Sunstein, 2008, p. 6), and then give as an example the publication by firms of “Toxic Release Inventories [...] which enable the media and environmental groups to produce an “environmental blacklist”” (p.191/193, quoted in Hausman and Welch, 2010). But this is effective because it makes pollution much more costly, so in fact it does significantly change the economic incentives (Hausman and Welch, 2010). The case of Nudges that work by making information more salient is less clear.

Let us take the example of the Ambient Orb: it was an orb provided by an electrical company that glowed red when energy consumption was high, and green when it was low. It reduced energy consumption by 40% (Thaler and Sunstein, 2008). The authors said that it was a Nudge because it made energy consumption more visual and transparent. Hausman and Welch (2010) argued that it was not a Nudge because it simply provided information. However, they ignore the fact that the Ambient Orb “changes the atmosphere of the choice architecture to work with people’s orienting *moods* (automatic thinking) in ways that reduce their energy use” (Selinger and Whyte, 2011). It is not merely the information, but the *color* that influences behavior by priming a specific mood. For example, the color “red” is intuitively associated with danger, and people would thus naturally want to act when the orb starts glowing red. Had the Ambient Orb merely provided numbers instead of using colors, it probably would not have worked so well.

In order to improve the clarity of the debate, several authors have come up with alternative definitions of a Nudge. Hausman and Welch (2010) state that “a nudge is any attempt at influencing choice in a predictable way without forbidding any previously available courses of actions or making alternatives appreciably more costly in terms of time, trouble, social sanctions, and so forth.” (p.15). Later, Hansen and Jespersen (2013) replace “choice” by “behavior” to account for the fact that a behavior is not necessarily the result of a choice. An alternative definition is provided by Hansen (2015): “A nudge is a function of (I) any attempt at influencing people’s judgment, choice or behavior in a predictable way, that is (1) made possible because of cognitive boundaries, biases, routines, and habits in individual and social decision-making posing barriers for people to perform rationally in their own self-declared interests, and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts.” (p.40) Most notably, in this definition, a nudge is no longer defined with respect to freedom of choice, incentives or information.

A problem with this definition lies with the second type of issues: why do nudges work? According to Hansen (2015), they take advantage of biases among other things, but some argue that the interpretation in terms of “biases” is wrong, and that Nudges work because of social intelligence rather than biases (Gigerenzer, 2015).

In order to avoid getting lost in endless theoretical debates, we chose to stay close to the original definition introduced in *Nudge*. Indeed, it remains the simplest and most practical of those definitions. Thus, in this work, we will use the following definition: “A Nudge is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or changing their economic incentives”. The only difference with the original definition is that we decide to not change economic incentives at all, while in the original definition, Thaler and Sunstein said “*significantly* changing their economic incentives”. This should make it much easier to differentiate between a Nudge and a simple financial reward. Also note that the definition does not make any assumption about whether it is rational or not to let oneself be “nudged”, it does not require one to wonder whether the change in choice architecture will have consequences on something as vague and open to interpretations as “trouble”, and it does not require one to evaluate whether the intervention is ethical or not. In the end, we believe that this definition is the least ambiguous of all, which makes it the best definition for a rigorous scientific approach.

Now that we have presented the Nudge approach and how we intend to overcome its most common criticisms, we will show how it can help with the issues we raised in the first part of this chapter.

Part 4: How can the Nudge approach help with the business problems we mentioned

Since the publication of the book in 2008, the Nudge approach has been very popular. For example, in 2009 Obama created the Nudge Unit, a team entirely dedicated to using the Nudge approach to solve behavioral issues in the US. The UK got the Behavioral Insight Team, created in 2010. Since then, many Nudges have been tested. Furthermore, people didn't wait for the publication of the book to apply the teachings of Behavioral Economics where the old solutions didn't work. After all, the denomination of Nudge can be applied retroactively to many different interventions.

A few of them are of particular relevance to our context. We already showed how it can help steer people toward the best retirement plan for their situation, but there are many more applications for the Nudge approach in the workplace. For example, Handgraaf, de Jeude and Appelt (2013) aimed at getting employees to save more energy at work. They compared a monetary

incentive (5 € reward for more eco-friendly behavior) with a social reward (a purely symbolic grade and a descriptive comment), and found that the latter was much more effective than the former, especially when it was made public. Holland, Aarts and Langendam (2006) showed that the mere preparation of an implementation plan (writing down in advance where, when and how someone intends to perform a specific behavior) significantly improved recycling behavior in a company.

There are many more examples like these, which means that the Nudge approach provides us with an extensive library of papers to draw ideas from. Therefore, while we will use the Nudge approach in new ways, we will still benefit from a lot of evidence to maximize our chances of success during our experiments.

When compared to monetary incentives, Nudges have a few major advantages. First, they are typically much cheaper to implement, which greatly limits the financial cost associated with trying one out and naturally increases their cost-effectiveness. Benartzi *et al* (2017) compared the cost-effectiveness of Nudges to that of monetary incentives with the same goal, and built the following table:

Table 1 – Cost-effectiveness of Nudges and monetary incentives

| Article | Intervention type | Treatment | Impact | Cost | Relative effectiveness |
|--------------------------------------|---|--|---|---|--|
| Allcott (2011) | Nudge | An independent company sent reports to residential consumers that contained comparisons to neighbors' electricity usage and tips for conservation. | 2.0% reduction in energy usage on average ^a | Approximately \$1 per report, with reports sent monthly, bimonthly, or quarterly | 27.3 kWh saved per \$1 spent ^a |
| Asensio & Delmas (2015) | Nudge | Researchers granted residential consumers access to a Web site sharing their detailed appliance-level electricity usage, with messages linking this usage either to health and environmental issues or to increased utility bills. | Health and environmental messages: 8.192% (4,306) reduction in energy usage; billing-oriented messages: negligible effect | \$3,019 per household | Health and environmental messages: 0.050 (0.026) kWh saved per \$1 spent; billing-oriented messages: negligible effect |
| Ito (2015) | Traditional (financial incentive) | Residents in California received discounts on their electricity bills if they reduced their summer energy usage by at least 20% relative to the previous summer. | 4.2% (1.3) reduction in energy usage in inland areas and negligible effect in coastal areas | \$3.70 per customer for rebates plus \$1.39 per customer for administrative and marketing costs | 3.41 kWh saved per \$1 spent ^a |
| Arimura, Li, Newell, & Palmer (2012) | Traditional (financial incentive and education) | Utility companies provided incentives and education to reduce energy usage during peak times and promote investments in efficient products. | 0.9% (0.5) reduction in energy usage during intervention period and 1.8% (1.1) reduction when including effects in future periods | \$10.83 per customer on average | 14.0 kWh saved per \$1 spent ^a |

Here we can see the Nudge tested by Allcott (2011) is much more cost-effective than both monetary incentives. One might also notice that the least effective of this treatment is not a monetary incentive, but another Nudge. The problem with what Asensio and Delmas (2015) tested is the very high cost of the intervention (\$3019 per household), which contradicts what we said in the previous paragraph. We deliberately refrained from including “cheap” in our definition of Nudging, because this is a subjective notion (the same expense could be perceived as expensive by one individual and cheap by another), but it is a key aspect of the approach that one should keep in mind when deciding what Nudge to test out, and we hope that this table illustrates why it is important. In our work, we will always make sure that what we test can be implemented at a low cost.

Second, to our knowledge, there is no evidence of a crowding-out effect when using Nudges. On the contrary, there is evidence that the effects of Nudges can linger even after the Nudge itself has been removed (Brandon *et al*, 2017). Cronqvist *et al* (2018) found that the effect of a Nudge to influence employees’ choice of pension plan lingered for up to two decades. It has also been shown that implementing the Nudge again after some time can reproduce the initial positive effects. For instance, Allcott and Rogers (2014) examined the long-term effects of the Opower Nudge, which consisted in sending a letter to certain households to let them know that their energy consumption was higher than their neighbors⁴. They noticed the Nudge could be re-implemented later on in the exact same way and yield the same results.⁵

This does not mean, however, that Nudges should replace monetary incentives. As we said earlier, there are many situations in which the latter are profitable. But, most importantly, Nudges can also work well with monetary incentives. For example, Hossain and List (2012) found that, by reframing pay bonuses as losses instead of gains (“you will be given a bonus, but you will lose it if you do not meet your objectives”), they could significantly increase productivity in a factory. This is a Nudge because it does not require changing anything about the actual incentives, simply modifying the way they are presented.

Because of this, we believe that the Nudge approach might be a good way to help alleviate some of the issues we mentioned earlier.

⁴ The letter was made visually appealing, it contained graphs showing how exactly their energy consumption compared to their neighbors. It also displayed an emoticon whose emotion changed whether the consumption was higher, lower or roughly equal. The letter also included tips to help them reduce their energy consumption. For more details on the experiment, see Allcott (2011).

⁵ They also noticed that the results in terms of energy consumption were only decaying at a rate of about 10-20% per year after the Nudge had ended, which means that the results do not disappear right after the Nudge is removed.

Conclusion

In this chapter, we have exposed the issues we want to tackle as well as the limits of monetary incentives. Then, we presented Behavioral Economics and one of its more popular applications, the Nudge approach. Afterwards, we discussed the main criticisms of the approach and how we aimed to overcome them in our work. We ended by showing why the Nudge approach is relevant to the issues we pointed out earlier, and why it is a viable alternative or complement to monetary incentives.

We hope that this short overview of the topic has successfully conveyed the potential of this approach, and why it is worth testing out in new areas. In the remainder of this thesis, we will present how we applied it to solve issues of survey participation, satisfaction measurement and motivation at work.

Chapter 2 - A Nudge is worth a thousand words: increasing web survey participation

Introduction

Web surveys have had an increase in popularity in the recent years (Couper and Miller, 2008). Unfortunately, response rate to web surveys remains 11% lower than response rate to other types of surveys (Manfreda *et al*, 2008). This is problematic as response rate is considered to be one of the major determinants of the quality of a survey (Ganassali, 2008), and if it is too low, it can generate nonresponse bias (Groves, 2006). While response quality consists of more than just response rate (see Ganassali, 2008 for a discussion of this notion), it is safe to assume that a better response rate will translate into better data for the researcher to analyze.

Researchers have come up with numerous ways to increase response rate to web surveys. Most of them focus on the design of e-mail invitations, on the use of incentives and reminders. The issue mainly lies in getting the e-mail addresses of potential respondents. To do so, we propose to use a channel that has not been studied much: phone. At the end of a phone survey we offered to the respondent who agrees the possibility to leave an email address so that he may be contacted to participate in future web opinion polls. In this paper, we will study how the nudge approach, which relies on behavioral economics, can help us to design the formulation of this request in a way that maximizes the rate of positive responses, and ultimately the participation rate.

In this paper, we will start by presenting the literature on web survey participation, before reminding the reader about the main points of the Nudge approach and showing how it relates to the current topic. Then, we will present the general design of our experiment. After that, we will present our two studies and their results. Finally, we will discuss the general findings that came from the whole research, and we will turn them into advices for decision-makers and researchers.

Part 1: Literature review

The literature on web survey participation

There is already an extensive literature about techniques to increase the response rates to web surveys (see (Fan & Yan, 2010) for a systematic review). Multiple contacts with the potential respondent (Manfreda *et al.*, 2008), in particular through reminders (Bruce, Hawkins, Sharp, & Keller, 2006; Cook, Heath, & Thompson, 2000), telling the potential respondent that he/she is part of a small

group (Pedersen & Nielsen, 2014; Porter & Whitcomb, 2003), personalization (Joinson & Reips, 2007) are the most reliable ways of increasing response rates. Incentives generally have a positive effect on response rate (Goritz, 2008; Heerwegh, 2006), yet their effect is occasionally found to be insignificant (Pedersen & Nielsen, 2014), which makes them unreliable.

In addition to this empirical literature, some papers attempted to explain what determines the decision whether to participate in a survey or not. According to leverage-saliency theory (Groves, Singer, & Corning, 2000) different individuals have different ways to evaluate the various aspects of participating in a survey (with factors such as incentive, topic and sponsorship), and the decision to participate to a survey is a function of how salient each aspect is made to the respondent and how much he values these different aspects. “The achieved influence of a particular feature is a function of how important it is to the potential respondent, whether its influence is positive or negative, and how salient it becomes to the sample person during the presentation of the survey request” (p.301) This theory has the advantage of explaining why the same feature may impact one respondent and not another. However, to rigorously apply it, one would need to first collect individual data in order to determine what features should be made the most salient to each person.

According to the planned-behavior approach to participation in web panel surveys, the intention to participate to survey is a function of moral obligations, attitude, subjective norm and perceived behavioral control (Bosnjak, Tuten, & Wittmann, 2005). In other words, a person decides to participate if they feel like it is their moral duty, if they have a favorable attitude toward this particular survey (which would depend on their personality, among other things), if they think that participating is the norm, and finally if they think that participating will not be too costly (this is what perceived behavioral control refers to: how time-consuming or tedious they perceive the survey to be). This theory is interesting but, just like the one we presented in the previous paragraph, some key variables are difficult to measure. For instance, it is very hard to assess whether the respondents feel morally obligated to answer, or how tedious they believe the process to be. And most importantly, it cannot provide any recommendation to improve participation rates, merely an explanation as to why certain people choose to participate while others choose not to.

Another theory proposes that the decision to participate depends mainly on peripheral aspects of the request, since people do not care much about that decision. As a result, they rely on heuristics to take a decision (Groves, Cialdini, & Couper, 1992). The authors cite 6 principles of persuasion: reciprocity, consistency, social validation, authority, scarcity and liking. This approach is more useful than the previous two for what we aim to achieve in this paper. Indeed, it is the most practical and does not require the assessment of other variables, such as moral obligation or the

personal saliency of a given feature, to provide recommendations for improving participation rate. Indeed, all there is to do is to modify the participation request in order to harness one of these 6 principles of persuasion. For example, one could emphasize that participating is the norm in the request (“most people we invite agree to participate”). This type of strategy is very similar to the Nudge approach. In the next subsection, we will remind the reader of its most important aspects and show why it is relevant in this particular context.

The Nudge approach

As we explained in the first chapter of this thesis, we chose to define Nudge as “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or changing their economic incentives”, which is very close to the original definition of Thaler and Sunstein (2008). Choice architecture is the way the information is presented. Thus, a Nudge changes behavior by simply reframing a choice, not by providing new relevant⁶ information or economic incentives, which is what makes it different from the way economists traditionally try to influence the behavior of economic agents.

It is easy to see how this approach relates to the social psychological theory of Groves, Cialdini and Couper (1992): in this context, Nudging is akin to changing the peripheral aspects of a request and, in fact, a lot of Nudges rely on social psychological findings. However, the Nudge approach is much richer and goes well beyond the 6 principles of persuasion. In this study, in addition to testing 5 of those principles, we also investigate order effects and various types of default framing effects (where an option is *framed* as the default choice). Furthermore, details are very important in Nudging, so we cannot simply test one implementation of a framing effect and consider the results to be definitive. There are many ways to harness one bias, and so we will test many different implementations of a given bias.

The real novelty of this research, however, lies in the medium. This is not the first paper which investigates the impact of framing on the decision to engage in a web survey. Many others have wondered (and rigorously tested) which e-mail subject line or body would elicit the most participation (see for instance Porter, 2012, Kaplowitz *et al.*, 2012). They varied both the content and the framing. However, to our knowledge, no other paper has investigated how framing effects fare in

⁶ We would like to insist on the word “relevant”. There are many interventions, including one in this work, that aim at changing behavior by showing people that their peers have already adopted the right behavior. This is indeed a new information, but one that would have absolutely no influence on the decision of a perfectly rational economic agent.

oral invitations by phone. In oral conversations, many other factors may be at play: the voice of the speaker, how well he speaks the language, the tone of voice, and so forth. With so many more interferences, will the framing still matter, or will it be drowned in the noise?

This particular context also gives our study another unique property: we are talking to people who already participated once, and just completed a survey to the end. This means that, in a sense, we are in a *foot-in-the-door* scenario. The foot-in-the-door effect refers to the fact that it's easier to convince people to do you a large favor if you already convinced them to do you a smaller favor right before, a phenomenon first shown by Freedman and Fraser (1966).⁷

This gives us the opportunity to investigate the effect of Nudges on a very particular sample. In most Nudge studies, it is impossible to determine how willing are people to perform a given behavior without directly asking them, which can introduce some biases (social desirability being the most obvious one: people will tend to give the answer that they think is most likely to please the interviewer (Edwards, 1957). Therefore, we can know whether different Nudges can have some form of effect on the population at large, but it's impossible to answer the question: what Nudges work best on people who have just shown a predisposition to perform the desired behavior? Answering this question, as well as determining what kind of Nudge yields the best results, is the purpose of the experiment we will present in the next section.

Part 2 - General study design

The process

This study was performed in collaboration with a French polling organization. This organization conducts weekly phone surveys on about a thousand persons in France. At the end of the phone survey, the interviewees are given the possibility to leave their e-mail address to participate later on. In this section, we will explain this process in detail.

Part 1 – The phone survey

⁷ What they did is actually not too far from the design of our study. In their paper, the “large request” was calling the subjects to ask them whether they agreed to have a team of 5 to 6 men come into their house. That team would list the household products they used, which should take about 2 hours, with the goal of writing a report for a public service publication. In half of the cases, the interviewer first called the subject a few days before making the large request and asked them if they were willing to take a small phone survey on the household products they used. When subjects did the phone survey prior to being called for the large request, they were twice as likely to accept it.

Step 1: The interviewer calls people who are selected via quota method. Most of the time, the calls are made between Friday evening and Saturday afternoon, and always by interviewers from a call center.

Step 2: The interviewer introduces themselves and asks whether the respondent wants to take a phone survey. If the respondent refuses, the call ends there – the interviewer does not ask anything else of them.

Step 3: If the respondent accepts, then the survey will be administered. The content of the survey will appear on a computer screen in front of the interviewer, who will then read it aloud to the respondent. They do read the survey word-by-word, with the utmost rigor, to ensure that there is no interviewer bias. The surveys typically last between 10 and 15 minutes, and are mainly about health, politics, news, personal habits, etc. They always include questions about sociodemographic variables, such as gender, income, or age.

Step 4: After the survey, the interviewer asks the respondent whether they want to participate to web surveys in the future⁸. This is where we introduce our test treatments. The formulation of these questions is written in advance and appears on the interviewer's screen after the survey has ended. Just as in step 3, questions are read word-by-word. In the original design, there was one specific formulation for this question, and it was the same for absolutely all phone surveys. This specific formulation will serve as the control treatment in our experiment.

During our experiment, the formulation that would appear on the screen was randomly selected among all treatments. Only people who declared that they used the internet at least once a day (a question that is always asked, along with sociodemographic questions) are enrolled in the study, which is about 80% of the sample interviewed. If a person refuses to answer or declares they use the internet less than once a day, then the computer selects the control treatment for the formulation.

If the respondent refuses to participate in future surveys, the call ends there.

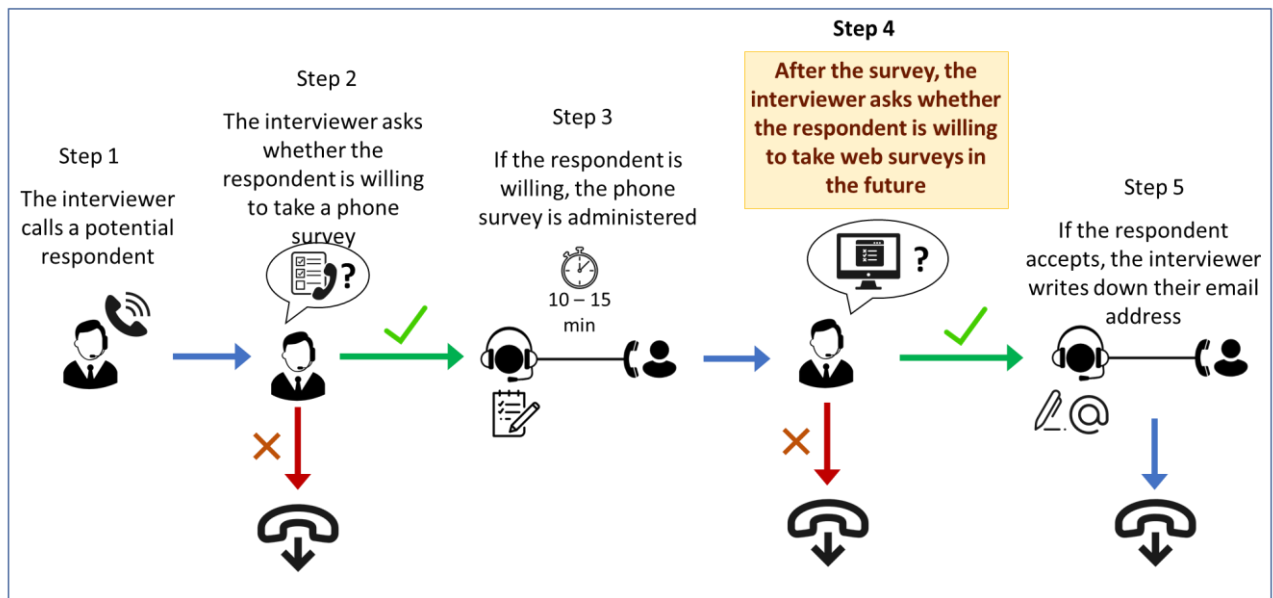
Step 5: If the person accepts to participate to web surveys, the interviewer will then write down their email address, and the call will end after that.

Note that there is no incentive to complete these phone surveys.

All 5 steps of this first part are summarized in the figure below:

⁸ In some cases, including in the original formulation that was in use before the experiment, the interviewer also asks the respondent whether they want to participate again to phone surveys. We did not include this detail in the figure in order to keep it as simple as possible and because this is not the focus of our study.

Figure 3 – The phone survey, first part of the process



Part 2 – The path to the web survey

Step 1: Within the next 24 hours, the respondent who left their email address will receive an email that asks them to register to the panel of web survey participants. To do so, they simply have to click a link in the email.

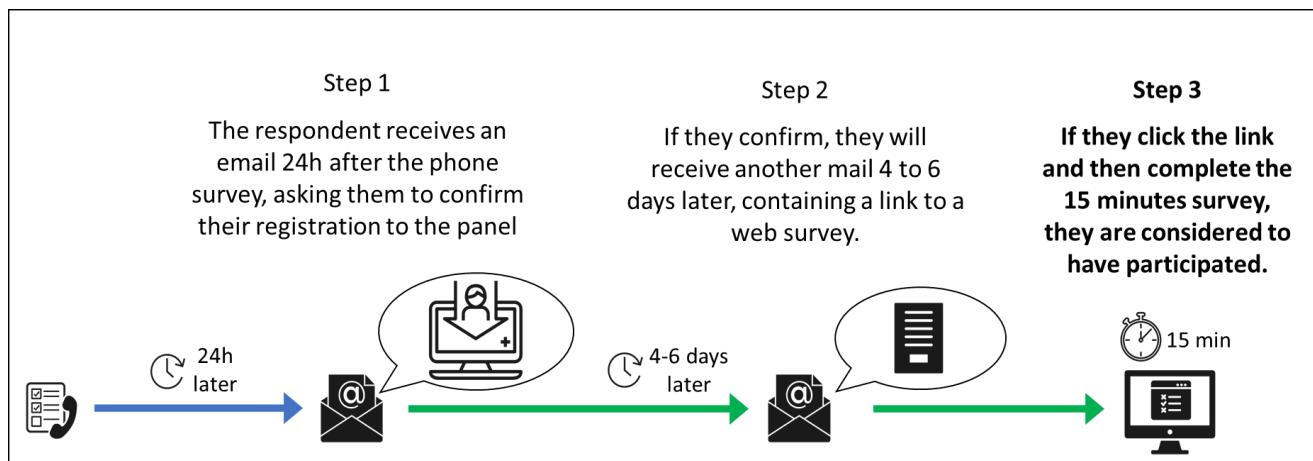
Step 2: Between 4 and 6 days later, and every week after that until they choose to unsubscribe, the respondent will receive an email that asks them to complete a web survey. The email will include a link to the webpage that contains the survey.

Step 3: The respondent completes the survey. It takes on average 15 minutes and, just like the phone survey, deals with politics, news, habits, etc., and there is no incentive to complete it.

At the end of the survey, the participant is thanked, and they may close the page. We only consider them to have participated if they went through the survey.

All 3 steps of this second part are summarized in the figure below:

Figure 4 – The path to the web survey, second part of the process



The experiments

The study consists of two separate experiments. Doing it in two separate experiments rather than testing all our treatments in a single larger experiment has two main benefits. First, by doing a second experiment at a different time of the year, we can spot a season effect – for instance, is one treatment more effective in Winter than in Summer? Interviewees are likely to be busier in Winter than they are in Summer, and so they may be less willing to be Nudged in the former than they are in the latter. Second, this allowed us to make adjustments for the second experiment based on what we learned while conducting and analyzing the results of the first experiment.

The sample and the process are identical, the only difference being the periods during which they have been conducted are the treatments that were tested.

The following three treatments are common across both studies:

Table 2 – Treatments implemented in both studies

| Treatment | Formulation of the participation request |
|------------|---|
| 0. Control | a. “Finally, would you agree to be contacted by email to participate in surveys?” b. “Finally, would you agree to be contacted again by phone for another poll?” |
| 1. Merging | “Finally, would you agree to be contacted by email to participate in surveys or by phone for another poll?” |
| 2. Polite | “Thank you for participating to this survey. If you wish to keep giving your opinion, would you agree to participate again by e-mail or by phone?” |

“Control” corresponds to the traditional way of inviting people to participate to a study: it is the formulation that was used by the firm that conducts these phone surveys before we did the

experiment. It is a neutral formulation that is devoid of any Nudge. Traditionally, people were first asked whether they wanted to participate by e-mail, and then, regardless of their answer, they were asked whether they wanted to participate by phone. Since this added more questions than necessary to a survey that might already be long, it was decided that both questions would be asked in a single sentence. This is why almost all of the “Nudge” formulations described below are in a single sentence. However, it is not impossible that this mere “merging” of two sentences into one already has an effect on participation rates, and if so, it would add a systematic confounding effect into the study that would make it impossible to attribute the results to the Nudges alone. Indeed, it would be impossible to differentiate between the effects of the Nudge and the effect of merging. This is why we added treatment 1: it is similar to treatment 0, with the only difference being that the two sentences are combined into one. This treatment will thus serve as a verification: if treatments 0 and 1 do not yield significantly different results in any way, then it means that all results seen regarding the remaining treatments can be attributed to the Nudge, and not to the merging.

Finally, the treatment labelled “Polite” aims to harness two effects. First, since expressing gratitude encourages prosocial behavior (Grant & Gino, 2010), we expect that by thanking them for their participation prior to inviting them, they will feel that their contribution is more valued, so they will be more inclined to participate benevolently again in the future. Second, by reminding them that they are already engaged into the process of giving their opinion, we hope to harness a foot-in-the-door effect. Finally, the overall formulation is friendlier than the two previous formulations.

We kept the same treatments 0 and 1 in both experiments because they serve as control groups. We also considered that keeping them exactly identical would allow us to detect any possible difference over time. For instance, if all Nudges in the second experiment lead to a better participation rate than the Nudges in the first experiment, but the control group also has a better participation rate, then it means the difference is to be attributed to time, and not to the formulations.

We also kept treatment 2 in both experiments as an additional form of control. Since this second experiment was performed at a different time of the year, in a different season, it is entirely possible that the results could change slightly. This is why we chose to keep one of the Nudge treatments as it is. We selected this one because it is simple and less likely to be perceived as coercive as some of the other treatments we tested.

Part 3 - Hypotheses

There are two variables that we are trying to increase with this research. The first one is the proportion of participants to the phone survey who agree to give their e-mail address to the interviewer. While our final goal is to improve future participation, this decision is the most immediate outcome of the treatment. There are at least five days between the decision taken at the end of the phone survey and the first web survey and, before that, the participant also has to officially register to the mailing list, which means that participation runs more chances of being influenced by unobservable factors. From now on, we will refer to this first variable as “acceptance rate”

The second variable is participation to at least one web survey. In order to be considered as having participated, an individual must have given his e-mail address, then accepted to register to web surveys, then answered a web survey invitation that would arrive in his mail box several days later, and then gone through the (approximately) 15 minutes that the survey will take. It is a heavy process, which is why we only want people to participate *at least once*: this is as far as the Nudge approach can get. There are many reasons why a respondent might quit after only a few surveys. For instance, he might have disliked the topics of the first surveys he filled, or simply be discouraged by the lack of incentives. Increasing sustained participation over the course of multiple web surveys is beyond the scope of what a simple formulation change at the end of a phone survey can do. All we can aim to achieve is to get them to try one survey to the end. From now on, we will refer to this second variable as “participation rate”.

The hypotheses are the same for both experiments and can be summarized in the following way: we expect all treatments except treatment 1 to result in a better acceptance rate and a better participation rate than the control group.

Now that we have explained our general hypothesis and method, we will move on to the first experiment.

Part 4 - Experiment 1

Method

This first experiment started on June 2015 and ended on September 2015. During this first experiment, we implemented the three treatments described above as well as 5 additional treatments.

Table 3 –Treatments of the first experiment

| Treatment | Formulation of the participation request |
|-------------------|--|
| 0. Control | a. “Finally, would you agree to be contacted by email to participate in surveys?” b. “Finally, would you agree to be contacted again by phone for another poll?” |
| 1. Merging | “Finally, would you agree to be contacted by email to participate in surveys or by phone for another poll?” |
| 2. Polite | “Thank you for participating to this survey. If you wish to keep giving your opinion, would you agree to participate again by e-mail to answer surveys or by phone for another poll?” |
| 3. Default | “Finally, if you wish to participate in other studies, please give us your email address.” |
| 4. Choice | “Finally, if you wish to participate in surveys or another poll, do you prefer to do it by e-mail or by phone?” |
| 5. Polite choice | “Thank you for participating to this survey. If you wish to keep giving your opinion, do you prefer to do it by e-mail or by phone?” |
| 6. Polite default | “Thank you for participating to this survey. If you wish to keep giving your opinion, I invite you to participate in surveys by e-mail. Please give us your email address.” |
| 7. Salience | “Did you find this survey: 1. Very interesting 2. Rather interesting 3. Interesting 4. Mildly interesting 5. Uninteresting <i>[If the answer is 4 or better]</i> In that case, I invite you to participate in surveys by e-mail. Please give us your email address.” |

For this first experiment, we focused on the well-researched default effect. The default effect refers to the fact that everything else being equal, people tend to stick to the default option. Notable examples of successful use of this effect are found in organ donations (Johnson and Goldstein, 2003), savings (Thaler and Benartzi, 2004) and environmental issues (Pichert and Katsikopoulos, 2008). Default options can be defined as options that will be selected if the chooser does nothing (Thaler

and Sunstein, 2008). In this experiment, registering them by default was impossible since the respondents must voluntarily provide their email address. Nevertheless, we wanted to see if we could harness a default effect merely by framing something as the default option. This could work because some evidence suggests that the default effect results from sensitivity to social norm: people choose the default option because the fact that it's the default suggests that it is the recommended option, a form of norm (Brown and Krishna, 2004).

We had two different ways of using the default effect. The first one was to make e-mail participation itself seem like the default option. In treatment 3, "default", instead of asking the respondent if he *wishes to provide* his e-mail address, he is directly asked to *provide* it. This treatment makes it seem that the standard behavior is to just provide one's email address for future participation.

The second way to use the default effect was to simply ask the respondent about their preference between phone and e-mail for future participation. This implies that participating again is just natural, and that the only matter that needs to be settled is how. Thus, participation to studies is the default option. We test this in treatment 4.⁹

Treatments 5 and 6 were added to test whether there would be an interaction between the default effect and the "polite" treatment. Finally, for treatment 7, we made the assumptions that the people who participated in such phone surveys actually enjoyed it, but may eventually get tired at the end of the interview. As a result, that enjoyment wouldn't be salient at that point, so they wouldn't be inclined to participate again anytime soon. By first asking them about their interest in the survey, we wanted to make the positive aspect of the experience more salient, so that they would be more inclined to participate again in order to relive that experience. Furthermore, by constructing the answers in such a manner that 4 out of 5 possible answers are positive, and by following with "in that case", we hoped to harness cognitive dissonance (Festinger and Carlsmith, 1959): if they found it interesting, then to be consistent with what they just declared, they should participate again.

The results of this experience are presented in the following subsection.

⁹ There is another peculiarity to this treatment: it is about conditional participation. In other treatments so far, people are asked to provide their e-mail address *if they wish to participate*. In other words, the fact that they give their e-mail address is a way of expressing an intention. Whether it is genuine or not is another matter entirely, but at least the implication is clear. In treatment 4, however, we ask them what contact mode they would prefer *if they wanted to participate*. Thus, giving the e-mail address in this treatment is less of a commitment to participating. It is impossible to determine in theory how much this will influence people's behavior, which makes this treatment all the more interesting to test.

Results

A total of 5549 persons were interviewed for the first experiment. 23% answered the phone survey on their mobile phone, while the others answered on a landline phone. There were 51% of males. The mean age was 45 years old. Age and level of education follow a normal distribution, while income follows a lognormal distribution, as could be expected given that the sample was selected specifically to be representative of the French population as a whole.

Acceptance rate

In the following table, we can see the acceptance rate for each treatment:

Table 4 – Acceptance rate per treatment

| Treatment | Acceptance rate (%) |
|-------------------|---------------------|
| 0. Control | 23.8 |
| 1. Merging | 25.3 |
| 2. Polite | 25.7 |
| 3. Default | 28.5** |
| 4. Choice | 34.3*** |
| 5. Polite choice | 31.4*** |
| 6. Polite default | 30.2*** |
| 7. Saliency | 26.9 |

Chi-Square tests of difference between treatment X and treatment 0. * : p-value < 0.10, ** : p-value < 0.05, *** : p-value < 0.01

In order to verify our hypotheses, we built a Logit model where the probability of accepting to leave one's email address is predicted by treatment, age and gender. The formula of the model is the following:

$$\Pr[\text{acceptance}=1] = F(b_1\text{Treatment} + b_2\text{Age} + b_3\text{Gender} + b_4\text{Income} + b_5\text{Weekday} + b_6\text{Education})$$

In addition to the treatment, we added a series of control variables that could impact the result. We expect older people to be less familiar with the internet and to prefer other methods such as the phone, which means they should be less likely to give an e-mail address (or to have one to begin with) and to participate. We could also expect that the weekday variable, which reflects whether the individual was interviewed on Friday or Saturday, could also have an impact: we believe that many people would be less busy during a Sunday afternoon than during a Friday evening. Finally,

we also investigated the effect of gender, income¹⁰ and education because they are major socio-demographic elements that could impact people’s decisions.

The results of the regression are shown in the table below¹¹. The reference group for the treatment variable is the control group, called “treatment 0” in our data.

Table 5 - Probability of acceptance

| <i>Predictors</i> | acceptance |
|-------------------|--------------------|
| | <i>Odds Ratios</i> |
| (Intercept) | 0.28 *** |
| 1.Merging | 1.11 |
| 2.Polite | 1.09 |
| 3.Default | 1.30 ** |
| 4.Choice | 1.71 *** |
| 5.Polite choice | 1.45 *** |
| 6.Polite default | 1.48 *** |
| 7.Saliency | 1.13 |
| age | 1.00 |
| female | 1.05 |
| income bracket | 1.03 ** |
| No diploma | 0.59 ** |
| BEPC | 1.29 * |
| CAP/BEP | 1.14 |
| BAC | 1.01 |
| saturday | 1.09 |
| Observations | 4642 |

* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

¹⁰ Respondents are not directly asked what their income is. Rather, they must say in what bracket the total income of their household is. The interviewer will ask them whether the income is less than 499€, or between 500 and 749 750 and 999, and so on until the respondent answers “yes”. To account for the fact that the distribution of income is usually log-normal, brackets on the higher end are larger than those on the lower end. For instance, they are asked whether their income is between 500 and 749 at first, and they are asked whether it is between 6000 and 7499 if they keep answering “no”. As a result of this methodology, income is an ordered variable, with the exception of the maximum value, 13, which simply refers to the people who refused to answer this question (see note 3 below for how we deal with this situation). In order to avoid clutter in the regression table, income is treated as a numerical variable, hence the “as numeric” indication in the table. We did another regression where it remained an ordered variable, and the results were not different in any way.

¹¹ One may note that the sample for these regressions is smaller than the one we announced at the beginning of this section by about 1000 people. This is approximately the number of people who refused to report their income to the interviewer, and had to be excluded from the regression. Prior to this model, we performed a simpler regression on the whole sample. The results regarding treatments effects were exactly the same as those we are about to present.

We used Odds-ratio to make the interpretation of the results more intuitive. An odds-ratio of 1 indicates no effect of the independent variable on the dependent variable. An odds-ratio significantly superior to 1 means that the independent variable significantly increases the odds that the dependent variable is equal to 1, while an odds-ratio significantly inferior to 1 means that the independent variable significantly decreases the odds that the dependent variable is equal to 1. The size of the coefficient can also be directly interpreted. For example, having an odds-ratio of 1.71 for treatment 4 means that the odds that a given person chooses to give us their email address when exposed to treatment 4 are 1.71 times the odds of them giving us their email address if they were exposed to treatment 0 (the baseline).

The first thing we can notice here is that the “merging” treatment, treatment 1, did not significantly impact probability of acceptance. This means that merging the two sentences did not significantly affect any variable of interest in this experiment, and so any result we find can be attributed to the Nudges.

We can see that treatments 4 to 6 had a significant impact on the acceptance rate, with treatment 3 also being very close to the 5% significance level. It would appear that framing participation as a default had a positive effect on acceptance probability.

Besides the treatment effects, we can also see that wealthier people were more likely to give their e-mail address, while people with no diploma were less likely to give their e-mail address than people who are at least at the BAC + 2 level¹², because their lower level of education might indicate that they are not that comfortable with writing. Both results go in the same direction: people with higher social status are more likely to give their e-mail address, everything else being equal. Furthermore, it is quite rare nowadays for someone in France to have no diploma at all, and so it is likely that such a person would simply not have an e-mail address to give, or feel uncomfortable to write and would rather just talk to someone. In addition, wealthier people are more likely to have easy access to a computer or a smartphone on which they could answer the web surveys. They are also more likely to be more comfortable with this mode of communication, which means answering a survey would take them less time.

Next, we looked at the participation rate.

Participation rate

The following table shows the participation rate per treatment.

¹² This is the reference level for the “education” variable

Table 6 – Participation rate per treatment

| Treatment | Participation rate (%) |
|-------------------|------------------------|
| 0. Control | 3.3 |
| 1. Merging | 2.1 |
| 2. Polite | 3.2 |
| 3. Default | 3.5 |
| 4. Choice | 2.7 |
| 5. Polite choice | 2.2 |
| 6. Polite default | 2.2 |
| 7. Saliency | 1.9 |

Chi-Square tests of difference between treatment X and treatment 0. * : p-value < 0.10, ** : p-value < 0.05, *** : p-value < 0.01

We attract the reader’s attention on the fact that the participation rates are overall very small. Participation rates vary between 2% and 3% between treatments. These low rates are the result of the heavy requirements that a person must go through to participate to at least one web survey, as we detailed in the previous sections.

To predict the probability of participating, we used the same model as before, changing only the dependent variable:

$$\Pr[\text{participation}=1] = F (b1\text{Treatment} + b2\text{Age} + b3\text{Gender} + b4\text{Income} + b5\text{Weekday} + b6\text{Education})$$

The results are displayed in the table below:

Table 7 - Probability of participation

| <i>Predictors</i> | participated |
|-------------------|---------------------|
| | <i>Odds Ratios</i> |
| (Intercept) | 0.01 *** |
| 1.Merging | 0.67 |
| 2.Polite | 0.89 |
| 3.Default | 1.02 |
| 4.Choice | 0.79 |
| 5.Polite choice | 0.65 |
| 6.Polite default | 0.63 |
| 7.Salience | 0.57 |
| age | 1.02 *** |
| female | 1.10 |
| income bracket | 1.08 ** |
| No diploma | 0.33 |
| BEPC | 0.54 |
| CAP/BEP | 1.06 |
| BAC | 0.70 |
| saturday | 0.95 |
| Observations | 4642 |

* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

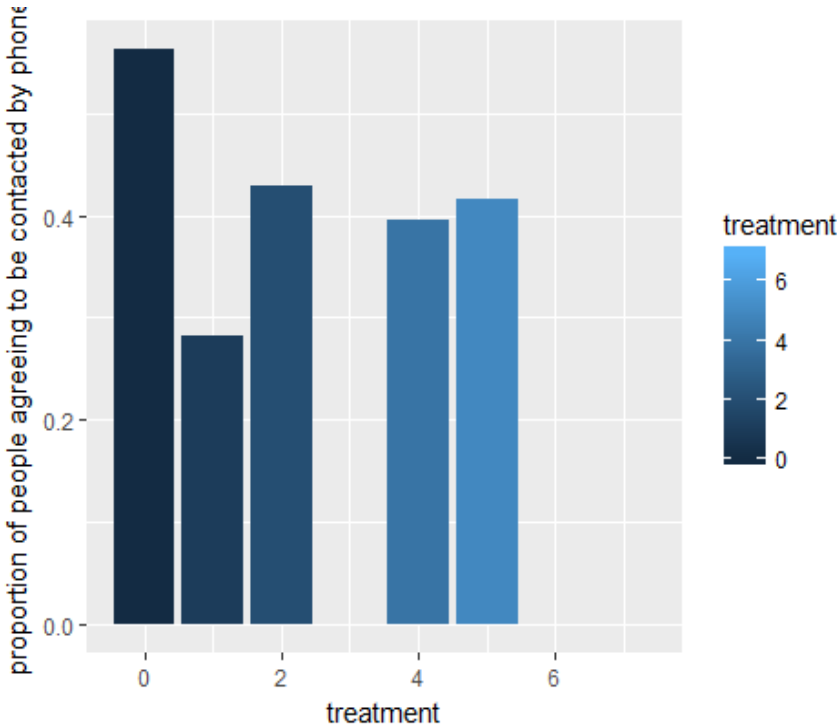
We can see that none of the treatments significantly improve participation rate compared to the control group. While Nudges did affect acceptance rate, the results did not transfer to participation rate, and we failed to influence it in this first experiment.

We can also notice that, contrary to our expectations, age significantly *increases* the probability of participating: older people are more likely to engage in web surveys. An effect all the more interesting when we remember that there was no age effect on acceptance rate. We also see the same income effect as observed on acceptance rate, but no effect of education. The age effect may simply be that older people would have more time (and patience) for the surveys that are proposed to them. Our interpretation of the income effect remains the same: wealthier people are more likely to have access to a computer, phone or tablet to answer the survey. They know it, which is why they are more likely to accept and, of course, they are more likely to participate since doing it is more convenient for them than for others.

Further analysis

After performing this analysis, we also checked the proportion of people who agreed to be contacted on the phone again. The goal of the study was to improve participation to web surveys, and it did not matter whether this was at the cost of participation to the phone survey or not. Nevertheless, it might be interesting to at least be aware of any side-effect. The table below shows the proportion of individuals accepting to be contacted on the phone again on each treatment.

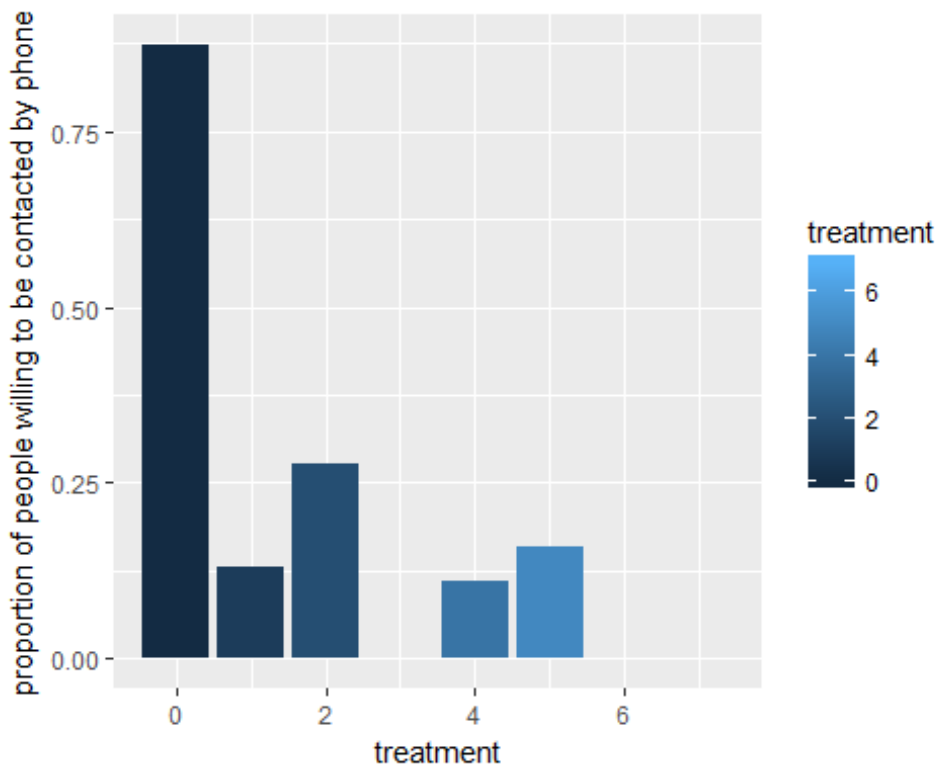
Figure 5 – Proportion of people agreeing to be contacted by phone again for future surveys



First, in treatment 3, 6 and 7, people were not asked whether they wanted to participate by phone again, which is why the proportion is 0 in those cases. Second, we can see a sizable difference between the treatments 0 and 1. While merging the two formulations did not have any significant impact on participation rate or the proportion of people leaving an e-mail address, it did have a significant impact on the proportion of people accepting to be contacted by phone again, as confirmed by a Chi-Square test (p -value < 0.0001). In fact, all differences of proportion between the control group and the test groups are significant.

To investigate this phenomenon further, we then looked at the subsample of people accepting to give their e-mail address: how many of them also agree to participate by phone? The answer is shown in the figure below:

Figure 6 - Proportion of people agreeing to be contacted by phone again for future surveys among those who gave an e-mail address



We can see that the proportion of people leaving an e-mail address who also agree to participate by phone is vastly superior in the control group compared to the other treatments. This means that this is not an effect of Nudging, but simply of merging the two formulations together. Indeed, having both elements in the same sentence implicitly make it sound like people who want to participate again may do so by e-mail or by phone, but that there is no point in doing it via both media. On the other hand, when the question was asked separately for each medium, it was not a matter of picking one of them, it was just a matter of deciding first if one wanted to participate by e-mail, second if one wanted to participate by phone. Two separate decisions.

This shows that one must be mindful of even the tiniest change in the way questions are framed, as even when one does not intend to include any Nudge, a sizable effect may occur.

Finally, we also looked at the proportion of valid e-mail addresses among all those given by the subjects, and we also looked at the proportion of people registering to the mailing list. Neither of those investigations yielded anything: the proportion of valid e-mail addresses is stable among all treatments (including the control), and the proportion of people registering to the mailing list is

always proportional to the participation rate, oscillating between 70% and 90% between treatments.¹³

This concludes our work in this first experiment. Then, we used what we learned here as a basis for selecting the treatments of the next experiment, and also to make a few changes to the design that we will discuss in the next section.

Part 5 - Experiment 2

Method

This second experiment started on October 2015 and ended on January 2016. During this second experiment, we tested the following treatments

Table 8 –Treatments tested in experiment 2

| Treatment | Formulation of the participation request |
|------------------|---|
| 0. Control | a. “Finally, would you agree to be contacted by email to participate in surveys?” b. “Finally, would you agree to be contacted again by phone for another poll?” |
| 1. Merging | “Finally, would you agree to be contacted by email to participate in surveys or by phone for another poll?” |
| 2. Polite | “Thank you for participating to this survey. If you wish to keep giving your opinion, would you agree to participate again by e-mail to answer surveys or by phone for another poll?” |
| 3. Thanks | “Thank you for participating to this survey. Would you agree to participate again by e-mail to answer surveys or by phone for another poll?” |
| 4. Consistency | “If you wish to keep giving your opinion, would you agree to participate again by e-mail to answer surveys or by phone for another poll?” |
| 5. Thanks direct | “Thank you for participating to this survey. If you wish to participate in other studies, please give us your email address. “ |
| 6. Norm | “Thank you for participating to this survey. We invite you to join the many French people who participate to web surveys. To do so, please give us your e-mail address. “ |
| 7. Ego | “Thank you for participating to this survey. You have been selected to join our privileged contacts, and thus participate to future e-mail surveys. To do so, please give us your e-mail address. “ |

¹³ This variation may seem large, but Chi-Square tests have revealed no significant differences.

As explained before, we used the same treatments 0, 1 and 2. With treatment 2, we tried to harness two effects at once, but perhaps we should have settled for something simpler. This is why we wanted to decompose this Nudge into two simpler formulations, each harnessing one of the effects we used. And because of a possible effect of time, we did not want to compare these two new formulations with the result we had in the previous section, so we had to reproduce the exact same Nudge in this experiment too to make sure it was comparable. The results of this decomposition are treatments 3 and 4. In treatment 3, we simply harness a gratitude effect, while in the section 4, we focus on the habit effect.

Treatment 5 is another way to harness a default effect along with a gratitude effect. Like before, we wanted to see if adding a simple “thank you” would make the default effect more effective, but this time we wanted to do it in a more straightforward manner.

Treatment 6 is inspired by another well-known behavioral effect: social norm. This is the fact that people tend to conform to what their peers do (Cialdini, Kallgren and Reno, 1991). In this treatment, we remind people that many people are already participating, which implicitly suggests that they should participate as well.¹⁴

In the last treatment, we test, in our specific context, an effect that is well-studied and generally powerful in survey recruiting: informing people that they have been specially selected to participate to the surveys (Pedersen & Nielsen, 2014; Porter & Whitcomb, 2003).

Having seen the surprising results that we had in the previous section regarding the decision to participate in future phone surveys, we asked that the interviewers always invite people to participate to future phone survey after they gave their decision regarding e-mail survey. We hoped that this would allow us to better investigate the possible unwanted effect of our Nudges on phone survey participation.

Results

A total of 5897 persons were interviewed for the second experiment. About 30% of them were interviewed on their mobile phone. There were 51% of males, and the mean age was 44 years old. Once again, age, income, and level of education follow a normal distribution.

¹⁴ One might think that with participation rates of around 3%, this formulation is a lie. However, keep in mind that polls are conducted on about a thousand individuals every week. A 3% participation rate means that there are 30 new participants *per week*. After a little more than 2 months, it already means there are 300 people participating, which is hardly negligible.

Acceptance rate

In the following table, we can see the acceptance rate for each treatment:

Table 9 – Acceptance rate per treatment

| Treatment | Acceptance rate (%) |
|------------------|---------------------|
| 0. Control | 19.1 |
| 1. Merging | 23.3 |
| 2. Polite | 23.1 |
| 3. Thanks | 24.5* |
| 4. Consistency | 27.1*** |
| 5. Thanks direct | 25.9** |
| 6. Norm | 24.8** |
| 7. Ego | 28.6*** |

Chi-Square tests of difference between treatment X and treatment 0. * : p-value < 0.10, ** : p-value < 0.05, *** : p-value < 0.01

We used the same model as in the first experiment to predict acceptance rate. The results are shown in the table below:

Table 10 - Probability of acceptance

| <i>Predictors</i> | acceptance <i>Odds Ratios</i> |
|-------------------|---|
| (Intercept) | 0.26 *** |
| 1.Merging | 1.23 |
| 2.Polite | 1.15 |
| 3.Thanks | 1.25 * |
| 4.Consistency | 1.52 *** |
| 5.Thanks direct | 1.38 ** |
| 6.Norm | 1.32 ** |
| 7.Ego | 1.55 *** |
| age | 1.00 |
| female | 0.88 * |
| income bracket | 1.03 * |
| No diploma | 0.57 *** |
| BEPC | 1.06 |
| CAP/BEP | 0.92 |
| BAC | 0.95 |
| saturday | 0.98 |
| Observations | 4909 |

* $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

First, there was still no significant difference between the control group and the “merging” group. Second, we can see that more Nudges have been successful in increasing acceptance rate than in the previous experiment, with treatments 4 and 7 being quite effective.

Just like in the previous experiment, we can see that people with no diploma are less likely to give their e-mail address, probably for the same reason we mentioned before.

Participation rate

In the following table, we can see the participation rate for each treatment:

Table 11 – Participation rate per treatment

| Treatment | Participation rate (%) |
|------------------|------------------------|
| 0. Control | 2.5 |
| 1. Merging | 3.1 |
| 2. Polite | 4.7* |
| 3. Thanks | 4.5* |
| 4. Consistency | 4* |
| 5. Thanks direct | 2.6 |
| 6. Norm | 2.2 |
| 7. Ego | 3.5 |

Chi-Square tests of difference between treatment X and treatment 0. * : p-value < 0.10, ** : p-value < 0.05, *** : p-value < 0.01

The results of the Logit model are shown in the table below:

Table 12 - Probability of participation

| Predictors | participated |
|-----------------|--------------|
| | Odds Ratios |
| (Intercept) | 0.01 *** |
| 1.Merging | 1.14 |
| 2.Polite | 1.71 * |
| 3.Thanks | 1.67 * |
| 4.Consistency | 1.64 |
| 5.Thanks direct | 0.99 |
| 6.Norm | 0.87 |
| 7.Ego | 1.28 |
| age | 1.02 *** |
| female | 0.79 |
| income bracket | 1.04 |
| No diploma | 0.38 * |
| BEPC | 1.20 |
| CAP/BEP | 0.48 *** |
| BAC | 0.87 |
| saturday | 1.06 |
| Observations | 4909 |

* p<0.1 ** p<0.05 *** p<0.01

We can see that, once again, older people are more likely to participate. People with a CAP/BEP are significantly less likely to participate than people with at least a BAC + 2, a result consistent with the general effect of education that we observed.

In this model, we can see that treatments 2 and 3 were effective at improving participation at the 10% significance level, with all others treatments failing to do so¹⁵. The common theme between these three treatments (including treatment 4, which a Chi-Square proved to be significant) is a friendlier and simpler tone. There is a very clear reason why, in our particular context, being straightforward and polite works better than more involved formulations. Here, we are talking to people who just proved that they are already willing to answer surveys. Hence, they don't need to be convinced that it's the norm, they already believe it, and they are already willing to help. Therefore, all we have to do here is to be polite, respectful, and to the point, and they'll gladly accept. This result would have probably been different had we talked to people who did not just agree to participate to a phone survey, which clearly shows the importance of context when using Nudges. Context is what makes these Nudges work here, and it is also why we do not find the same results as in the first experiment regarding treatment 2, even though it is exactly the same, as we will see in the next section.

Differences between experiments 1 and 2

The table below shows the differences in acceptance rate and participation rate between the two experiments. We chose to only include the treatments that the two experiments have in common, as it would make no sense to compare the results for other treatments.

Table 13 – Acceptance rate and participation rate in both experiments

| Treatment | Acceptance (%) | | Participation (%) | |
|------------|----------------|--------------|-------------------|--------------|
| | Experiment 1 | Experiment 2 | Experiment 1 | Experiment 2 |
| 0. Control | 23.8 | 19.1 | 3.3 | 2.5 |
| 1. Merging | 25.3 | 23.3 | 2.1 | 3.1 |
| 2. Polite | 25.7 | 23.1 | 3.2 | 4.7* |

*: Significantly different from treatment 0 as indicated by a Chi-Square test

¹⁵ One might note that treatment 4 is insignificant in the model, but significant in the Chi-Square tests. This is because the sample of the model is actually smaller, as it does not include people who refused to give their income bracket. Nevertheless, the p-value for treatment 4 is extremely close to the 10% significance level. This, combined with the results from the Chi-square tests, mean that we can consider it to be successful at improving participation rate.

We can see that the acceptance rate is consistently lower in the second experiment. Chi-Square tests of differences have revealed that acceptance is significantly lower in experiment 2 (p-value = 0.01411), while participation is significantly higher in experiment 2 at the 10% level (p-value = 0.05858), an effect driven by the difference in participation regarding treatment 2 (p-value = 0.06284). There is a simple explanation for this.

The main difference between experiments 1 and 2 is that the former took place during summer while the latter took place during winter. In summer, it is likely that people will be less busy and more relaxed. With such favorable circumstances, they'll gladly accept to do a favor to the interviewer by agreeing to also participate in web surveys. The drawback is that, because of the hot-cold empathy gap¹⁶, they will overestimate their future willingness to participate. Later on, when they get back from holidays, they won't be in such a good mood anymore and will not necessarily want to take these surveys when they make their way to their digital mail box.

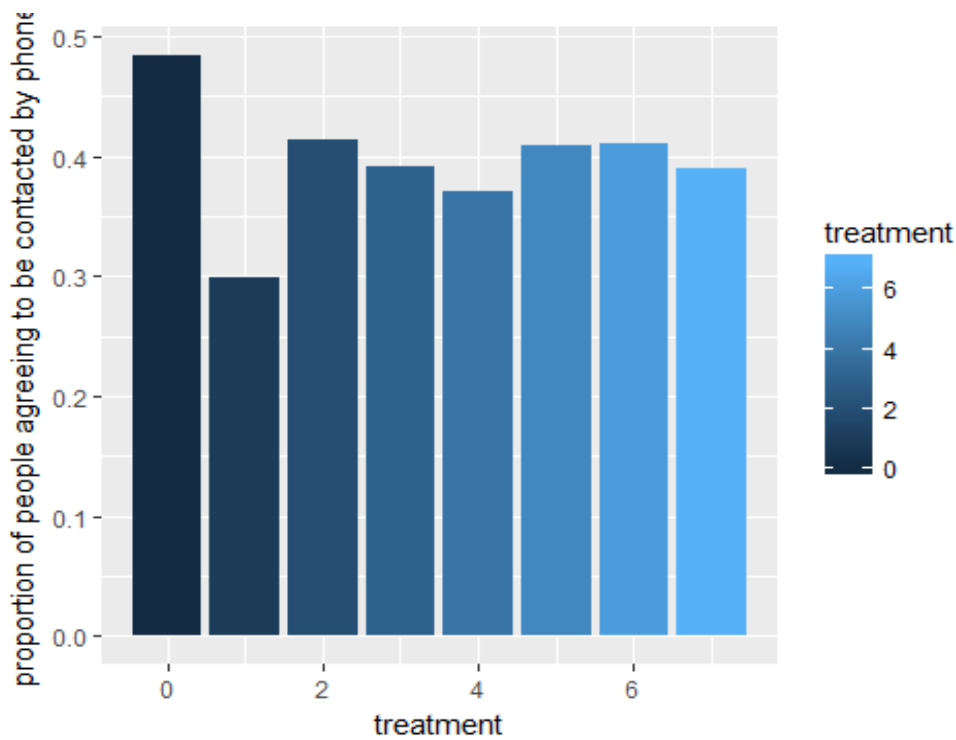
In winter, they are likely to have the opposite frame of mind: they will be very busy, and thus less likely to be willing to give their e-mail address to the interviewer for future web surveys. However, if they do, it is definitely not because they overestimate their ability or willingness to answer later on, which is why the participation rate is generally unfazed by the change in acceptance rate. And because of this lack of time, it is understandable that the simpler and more straightforward formulations are better appreciated than they were in summer. That change is enough for a previously ineffective Nudge to become effective.

Further analysis

We then decided to check the proportion of people agreeing to participate by phone again, in order to see if the pattern was the same as in the previous experiment. This time, people were asked whether they wanted to participate by phone in every single treatment, in order to provide more data for us to analyze. The figure below shows the proportion of people agreeing to participate by phone in each treatment.

¹⁶ This is a cognitive bias that make people underestimate the influences of visceral drives on their own attitudes, preferences, and behaviors (Van Boven et al. 2013).

Figure 7 - Proportion of people agreeing to be contacted by phone again for future surveys



Once again, the control has a significantly higher proportion than every other treatment, confirming what we already observed in the previous experiment: two separate questions make people more likely to also participate by phone, while having only a single one makes them more likely to just pick one medium. However, we can see that the difference is noticeably smaller, a result of asking people who gave us an e-mail address whether they also wanted to participate by phone. This suggests that, indeed, the drop in phone participation was the result of the formulations mistakenly suggesting that people should just pick one medium, and not two.

Finally, we also checked the proportion of valid e-mail given, and we found no significant differences between treatments: it remains at a proportion of 70% across all the Nudges. The proportion of people who do participate after registering is also stable, albeit smaller than in experiment 1: it was between 70% and 90% in experiment 1, while it is between 60% and 80% for experiment 2. However, this difference proved to be insignificant.

Part 6 - Discussion

More than anything, this study showed how the tiniest change in the way a simple question is formulated can have a significant impact on choice and behavior. Many of the Nudges affected the

immediate decision of the respondent to disclose their e-mail address, which was already a significant step. But, further than that, two of them even managed to influence a behavior that occurred one week later. This means they did more than simply influence the immediate answer of the respondent: they made a lasting positive impression on him, enough that he actually wanted to participate, and not just give an e-mail address to please the interviewer. Another most unexpected way in which this fact is illustrated is by the effect of the “merging” treatment” on the decision to participate by phone again, which dropped from 80% in the control group to 15% in the “merging” group in the first experiment, and from 50% to 30% in the second one. The merging of the two sentences might seem trivial, and wasn’t even intended to have any effect at all since it was a mere verification. Yet our results prove that even a subtle change in frame can have a big effect.

Another important result is the disconnection between acceptance rate and participation rate. While increasing acceptance rate proved to be relatively easy, increasing participation was not, and the only treatments that significantly increased participation had little to no effect on acceptance. In fact, the treatments that worked best for participation were the simplest ones, while the most heavy-handed ones were ineffective. This shows the importance of simplicity in Nudging. It also means that because someone is willing to give their e-mail address does not mean they intend to participate. As such, we should probably not insist on getting contact information at the expense of anything else if the ultimate goal is to increase participation rates. In the first experiment in particular, many treatments were made essentially to get this address, and not particularly to generate engagement, and we saw that while they did allow us to get more e-mail addresses, it did nothing on participation rate.

We could have thought that in certain treatments, people would just give a fake e-mail address to get rid of the interviewer, but this was not the case. The implication is that if someone wants to collect valid e-mail addresses but does not care about making people participate to web surveys, this study shows many good ways to do it.

The global difference in acceptance rate and in the effectiveness of one Nudge between experiments 1 and 2 can shed some light on what happened in this study. There is one single, major difference between these experiments: the former took place in Summer, while the second took place in Winter. As we already explained in the previous section, in Summer, most people will be on vacation, and we can expect them to be in a better mood and also to overestimate their capability and willingness to answer later on, when their mood won’t be as good. However, the end of the year tends to be a busy period, and people will be less willing to take time out of their day to answer web surveys. Indeed, since they will be busier, it will mean that time is more precious to them at that

point. To make the following explanations easier to follow, we included the formulations of experiment 2 again in the table below:

Table 14 - Treatments tested in experiment 2

| Treatment | Formulation of the participation request |
|------------------|---|
| 0. Control | a. "Finally, would you agree to be contacted by email to participate in surveys?" b. "Finally, would you agree to be contacted again by phone for another poll?" |
| 1. Merging | "Finally, would you agree to be contacted by email to participate in surveys or by phone for another poll?" |
| 2. Polite | "Thank you for participating to this survey. If you wish to keep giving your opinion, would you agree to participate again by e-mail to answer surveys or by phone for another poll?" |
| 3. Thanks | "Thank you for participating to this survey. Would you agree to participate again by e-mail to answer surveys or by phone for another poll?" |
| 4. Consistency | "If you wish to keep giving your opinion, would you agree to participate again by e-mail to answer surveys or by phone for another poll?" |
| 5. Thanks direct | "Thank you for participating to this survey. If you wish to participate in other studies, please give us your email address. " |
| 6. Norm | "Thank you for participating to this survey. We invite you to join the many French people who participate to web surveys. To do so, please give us your e-mail address. " |
| 7. Ego | "Thank you for participating to this survey. You have been selected to join our privileged contacts, and thus participate to future e-mail surveys. To do so, please give us your e-mail address. " |

Let us keep in mind that the two most effective treatments at that time were treatment 2, "polite", and treatment 3, "thanks", treatment 2 having been unsuccessfully tried in experiment 1. What do those two treatments have in common? People were simply thanked for their participation, and then asked in the most simple and friendly way possible, without any attempt at influencing

them, whether they wanted to participate to web surveys. In other ways, there was one important factor at play here, that mattered in Winter but not in Summer: recognition. By thanking them, we recognize that they did us a favor by participating without any incentive whatsoever. In Summer, this does not matter: their subjective cost of time is low, so they didn't "spend" much in order to help the interviewer out by accepting to take on the phone survey. In Winter, however, since they will be busier, their subjective cost of time will be higher, and so they will have to pay a higher "temporal cost" in order to take on the phone survey. Which means that this time, they would greatly appreciate this simple and universal recognition for their efforts, and perhaps even reciprocate by still participating later on.

This explanation raises another question: people were also thanked in treatments 5 to 7, so why were they not as effective? Because we inadvertently ruined the effect of recognition by adding what followed. Treatment 5 sounded more like a command than a humble request, which does work in order to get more e-mail addresses but does not necessarily put people in a good disposition (in addition to being borderline unethical). As for treatments 6 and 7, they were both relying on well-known behavioral levers, as explained earlier. While the literature we presented shows that they can still be effective, even today, they can still be a bit heavy-handed: they induce pressure to the respondent by implying that they *should* participate, while in treatments 2 and 3, the interviewer simply asks them if they *want to*.

In the end, there are a few simple lessons we can draw from this. The most obvious one would be that Summer is a good time for collecting e-mail addresses in such a way. The second one is that when people are busy, or when one wants to make a request that will incur a certain cost, the best way to do so is to simply be polite and recognize the effort that was made, rather than trying to implement one or many behavioral levers. The Nudge approach is very powerful, but it can also be overused, and one should know at what times one should stay simple. This is one of those times.

Future research

This study leaves a lot of room for further research. One could start by investigating in more detail this difference in treatment effectiveness over time. It might also be interesting to transfer our Nudges to different contexts, for example by testing them in e-mail invitations, or maybe in face-to-face interviews (when interviewers are in the street looking for persons willing to answer their survey). In this study, we observed a very specific sample of people: individuals who already agreed to participate to a phone survey all the way to the end. Would these Nudges work as well on people

who typically do not answer phone surveys? It is a question that is worth investigating. And, of course, we have barely scratched the surface of what could be done in the context of phone surveys, so one might want to try a whole new range of Nudges in addition to what we already tested. For instance, in this study we aimed to get people to participate many times in the future, even if we just measured one instance of participation. But what if one only wants them to participate once? There are many reasons for wanting such a thing, a greater diversity in respondents being the main one. Asking someone to “participate to web surveys” can seem daunting as it implies that they would have to take many surveys. But if we asked them to “simply participate to a single, short web survey”, the task would seem less intimidating, and it is more likely that they would do so.

Another problem that we could not adequately address was the possibility of a selection bias. Since people who chose to participate are necessarily people who chose to leave an e-mail address, one might wonder whether there was a selection bias. We consider that if there was, then the variables that heavily influence the decision to participate will be the same as those that heavily influence the decision to leave one’s email address. However, we showed that the results were different, especially in terms of treatment effects. Nevertheless, accounting for this possibility would have added precision to our second model. We did not do it because we failed to find a method that would work well enough for our case. The classic Heckman two-step model (1979) seemed inappropriate because our two models are essentially the same, and because of the fact that they are both Logit. Furthermore, while we had access to demographic variables, they are hardly the only elements that matter. Different people have different personalities, aspirations and opinions, and we could not measure such variables in our experiment, as it would have made it too heavy for the subjects. It would be very interesting to replicate this study while measuring these variables to see if, compared to the larger sample, people with specific personality traits or opinions are more likely to participate than others. Likewise, replicating the essential elements of the study, while slightly altering it to make it possible to conduct a precise investigation of a selection bias, could prove very interesting and could even reveal a major bias in classical survey methodologies.

Finally, we could not have access to enough data to tell whether people who participated kept taking web surveys later on. It is possible that certain Nudges made them want to participate just once, while others made them sufficiently motivated to keep participating in the long run. Investigating how Nudges can create different patterns of participation would be very interesting.

Conclusion

In this study, we presented the current problem with web surveys, namely their low participation rate, and offered a novel solution to it. We studied this solution in a context in which it was never applied before, phone surveys. Then, we analyzed its effect and we have shown that some Nudges did manage to improve participation rate. While the proportion of success and the size of their effect is small, this solution was implemented at almost no cost, since we only changed a couple of sentences in a script, and when it worked, it almost doubled the participation rate.

In the end, we hope that this study has given Nudge practitioners new ideas on where to apply this approach, and that it has provided those who conduct web surveys regularly new ways to improve participation rate.

Chapter 3 - Nudge and job satisfaction: the development of a new measure

Introduction

Working conditions have always been a major concern for modern businesses and public authorities. Going to work every morning could bring enjoyment to some people while being a source of anxiety to others. These two feelings that a person could experience are all but an anodyne phenomenon. Companies have realized it long ago and learned to apprehend the importance of job satisfaction by its resulting consequences, that could be either virtuous or disastrous on a whole person's life as much as on the company involved.

Knowing how important it is to understand the outcomes of job satisfaction and its implications on employee's psychological well-being as much as on the organization's efficiency, research has developed a large literature, digging into the understanding of the nature of this important variable, the way to assess it and its potential causes and consequences.

Job satisfaction is often described as *the degree to which people like their job* (P. Spector, 1997, p. vii). Knowing that this variable could have numerous causes and consequences, the modern management began to systematically assess employees' job satisfaction and consider it as a common and vividly recommended action in numerous companies.

Traditionally, job satisfaction is measured once every year via an extensive questionnaire. However, this method suffers from several limitations. Notably, satisfaction is subject to the "*peak-end effect*" (Kahneman *et al.*, 1997), which means that measuring it once a year does not guarantee an accurate representation of how satisfied the employee is throughout the year. Additionally, the answers to such a test fluctuate according to the mood of the subject when he is responding.

In this study, job satisfaction is investigated at its first stages: "The internship experience". In fact, among all years of studies, internship is maybe the most important stage to occupational integration. More than that, it is indisputable that internship experience represents a key learning period and an important preparation step for the future career. As said by (Gupta, Burns, & Schiferl, 2010, p. 28) "Internships provide direct business contact for students, usually in an employment setting". This is why expanding knowledge concerning factors that could induce higher satisfaction experience may also help educators as much as business to educate and recruit students successfully.

The purpose of this paper is to measure satisfaction in two steps: the first one periodically (every month) during the internship experience and the second one just at the ending of it. This choice is motivated by a double drive; unveil potential bias occurring in the measure of satisfaction as much as reaching the possibility to study more deeply some influences that could occur between satisfaction, learning and other related measures described below.

This chapter will be organized as follows. In the first part, we will present the literature on job satisfaction and its measurements, with a focus on internships. In the second part, we will present our hypotheses. In the third part, we will detail the design of the study. In the fourth part, we will analyze our results, and in the fifth part, we will discuss our findings, their implications, the limits of this study and directions for future research.

Part 1 – Theoretical background

Professional enjoyment and satisfaction are the most studied issues in organizational psychology. In fact, assessing employee’s job satisfaction has become a common routine in most companies showing concerns about the psychological well-being of their workforce. But how could we define professional of job satisfaction?

According to (P. Spector, 1997, p. 02), “*job satisfaction is simply how people feel about their jobs and different aspects of their jobs. It is the extent to which people like (satisfaction) or dislike (dissatisfaction) their jobs. As it is generally assessed, job satisfaction is an attitudinal variable...job satisfaction can be considered as a global feeling about the job or as a related constellation of attitudes about various aspects or facets of the job*”. From that state of facts, the table below shows nine of the most common facets used to assess job satisfaction.

Table 15 - Common facets of job satisfaction studies (P. Spector, 1997, p. 08)

| Facet | Description |
|-----------------------------|---|
| Pay | Satisfaction with pay and pay raises |
| Promotion | Satisfaction with promotion opportunities |
| Supervision | Satisfaction with the person’s immediate supervisor |
| Fringe benefits | Satisfaction with fringe benefits |
| Contingent rewards | Satisfaction with rewards (not necessarily monetary) given for good performance |
| Operating conditions | Satisfaction with rules and procedures |
| Coworkers | Satisfaction with coworkers |
| Nature of work | Satisfaction with the type of work done |
| Communication | Satisfaction with communication within the organization |

Internship or professional satisfaction?

Internship could be considered a key part of a complete college education. In fact, more than 90 percent of colleges propose internships or similar experiences to their students (Divine, Linrud, Miller, & Wilson, 2015). This fundamental period offers a first footstep in the business world where students have, maybe for the first time, the opportunity to apply the knowledge gained in classrooms for numerous years.

As stated by (Coco, 2000, p. 41), internships have the purpose “to provide a planned transition from the classroom to the job” and even if this earliest job experience could be in some cases paid and sometimes not, some empirical studies have found that students enrolled in internships received a lot more job offers than others (Di Lorenzo-Aiss & Mathisen, 1996; Gault, Redington, & Schlager, 2000).

This is why we made the choice to investigate professional satisfaction throughout the internship experience.

Assessing job satisfaction

Even if different methods are practiced by companies all over the world, measuring job satisfaction is still problematic. In fact, numerous studies have found different biases inherent to job satisfaction (Fredrickson & Kahneman, 1993), (Kahneman, 2000). Also, methods to measure job satisfaction are diverse: Single item measures, Herzberg's Two-Factor Theory, Minnesota satisfaction questionnaire, The 6-item *Global Job Satisfaction Survey* and others (Wanous, Reichers, & Hudy, 1997), (Maidani, 1991), (King, 1970). Below are some of the most common measures.

Classical measures

Being less time consuming and less expensive, questionnaires are usually preferred to interviews when measuring job satisfaction. For the purpose of seeking validity and reliability, many “standard” job satisfaction scales are generally preferred to assess employee satisfaction even if consulting firms tend to use their proper scales when conducting studies for others. These scales are generally customized according to the specific areas of interest claimed by a company.

Far from pretending to be exhaustive, we can mention four of the most popular job satisfaction scales used by academics:

- The Job Satisfaction Survey (JSS) developed by (P. E. Spector, 1985) assessing the nine facets listed above in table 13 using a 36 item questionnaire.
- The Job Descriptive Index (JDI) developed by (Smith, Kendall, & Hulin, 1969) is the most popular among organizational researchers. The JDI assesses five facet scores (work, pay, promotion, supervision, coworkers) using 72 items.
- The Minnesota Satisfaction Questionnaire (MSQ) (Weiss, Dawis, & England, 1967), is also popular in the research community. The scale can be used in two forms: a short version with 20 items or a longer one with 100 items covering 20 facets that permits to dig into very specific areas of job satisfaction facets that are listed in the table 2.

Table 16 - The twenty facets assessed in the Minnesota Satisfaction Questionnaire (MSQ)

| |
|--------------------------------|
| Activity |
| Independence |
| Variety |
| Social status |
| Supervision (human relations) |
| Supervision (technical) |
| Moral values |
| Security |
| Social service |
| Authority |
| Ability utilization |
| Company policies and practices |
| Compensation |
| Advancement |
| Responsibility |
| Creativity |
| Working conditions |
| Coworkers |
| Recognition |
| Achievement |

- Michigan Organizational Assessment Questionnaire Subscale developed by (Cammann, Fichman, Jenkins, & Klesh, 1979) is the shorter and simpler one, containing three-item overall satisfaction subscale shown in table 3. Each item has seven possible choices: “Strongly disagree”, “Disagree”, “Slightly disagree”, “Neither agree nor disagree”, “Slightly agree”, “Agree”, “Strongly agree”.

Table 17 - The three items of the Michigan Organizational Assessment Questionnaire Satisfaction Subscale

| |
|--|
| 1. All in all I am satisfied with my job |
| 2. In general, I don't like my job |
| 3. In general, I like working here |

Major critics

Being preferred due to the easy and simple way to measure it, the assessment of a unique general reported professional satisfaction at the end of a period has its downsides. In fact, economically speaking, the measure of reported satisfaction could be equated to what (Kahneman, Wakker, & Sarin, 1997) named, the measure of “*experienced utility*”.

Yet, on the basis of what Kahneman *et al.* tried to disentangle in the famous “Back to Bentham” paper, experienced utility should be considered by “*instant utility*”, that is defined as “*a measure of hedonic and affective experience, which can be derived from immediate reports of current subjective experience or from physiological indices. Instant utility corresponds to the dimension of “intensity” in the writings of Bentham, Jevons, and Edgeworth*”(Kahneman *et al.*, 1997, p. 376).

Transposed to our case of measuring a relatively long period outcome (satisfaction), this experienced utility should be evaluated in terms of “Temporally Extended Outcomes” TEO. Two forms of utility could be inferred from TEO’s; “*remembered utility*”, that is “*a measure on past TEOs, which is inferred from a subject’s retrospective reports of the total pleasure or displeasure associated with past outcomes*”(Kahneman *et al.*, 1997, p. 376). And total utility that is “*a measure on possible TEOs, which is constructed from temporal profiles of instant utility*” (Kahneman *et al.*, 1997, p. 376).

Assessing a unique retrospective reported satisfaction at the end of a relatively long period of job exercise (like a year as it is commonly practiced) could be considered as an outcome of a “remembered utility” measure. This latter is often accused of being potentially invalid because of numerous biases due to the temporality it involves. To overcome this obstacle, a promising alternative is offered by real-time measures of instant utility. In fact, repetitive real-time measures offer to “*avoid the biases of memory and evaluation that affect retrospective judgments of pleasure, pain, and well-being*”(Kahneman *et al.*, 1997, p. 394).

For example, one of the most documented biases raised by remembered utility assessments is the *Peak-End* effect¹⁷. It refers to the recall of the most intense or the last value when evaluating a remembered utility.

Emerging alternatives

Daniel Kahneman with his famous book “Thinking, Fast and Slow” (Kahneman, 2011), has largely contributed to vulgarize the existence of two types of mind, namely System 1 and System 2, that he defines as followed :

“System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control.

System 2 allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration” (Kahneman, 2011, p. 21).

More intuitively, system 1 could be assimilated to a “hot and immediate” thinking while system 2 represents “cold rationalization” thinking. We made the choice, in this study, to use a method that offers to capture system 1 thinking by taking several point measurements under the same conditions but at random times. In consequence, the satisfactions are then measured hot, in the moment (therefore, not deformed), rather than cold, rationalized and perhaps transformed. Finally, if the duration of the activity is quite long (here, about six months), these satisfactions distributed over time can be more easily related to the precise activities carried out at the same time.

This methodology is known as the ESM (Experience Sampling Method). In fact, as described by (Hektner, Schmidt, & Csikszentmihalyi, 2007), ESM has the *ability to capture daily life as it is directly perceived from one moment to the next, affording an opportunity to examine fluctuations in the stream of consciousness and the links between the external context and the contents of the mind.* In our case, we will test this alternative way of measuring satisfaction by recording satisfaction once a month, at a seemingly random time, with very short questionnaires¹⁸. We compare this new method to the traditional method, and then proceed to discuss the additional information that this allow us to gather.

Aimed at colleges and much more?

¹⁷ For more documented evidences, see (Schreiber & Kahneman, 1996) and (Clark and Georgellis, 2004).

¹⁸ Note that ESM are often much more involving, with subjects being interviewed 4 or 5 times *a day*. However, such a heavy method would probably have made it very difficult to find participants for this experiment.

Satisfaction during the internship being an important result that colleges as much as businesses should strive to fulfill, it has been observed that students with a more satisfying internship experience tend to have positive feeling toward both their academic institutions and the job search process and afterwards, when beginning their careers, are better able to contribute to their companies and respond to their needs (Paulins, 2008).

While firstly being concerned about the learning environment of students, our objective in this study is to shed a new light on how students view their first job experience, providing a punctual and repeated reporting of satisfaction and learning all along the internship experience. Moreover, our aim is to identify relevant factors playing role in ensuring a satisfactory experience from completing internship, our purpose being to draw some new implications for students and colleges based on our findings.

On the other hand, while being of crucial importance in the business sphere, we could say that, from a humanitarian point of view, professional satisfaction instinctively reflects fair and respectful treatment by the employer, occasioning in a balanced state of mind and emotional well-being. A complementary pragmatic and utilitarian perspective would say that, as a result of that psychological health, professional satisfaction could lead to productive behaviors within organizations. This is why each revealed source of satisfaction or dissatisfaction is enough to justify greater investigations of this variable.

Part 2 – Hypotheses

This study has a double objective. The first one is to develop an alternative measure of job satisfaction and evaluate its usefulness, while the second is to improve internships in general. This second objective encompasses the first of course, since a better measure of satisfaction will benefit interns¹⁹, but will require us to go deeper, by also observing additional variables that could affect satisfaction.

As such, we will split the hypotheses and the results sections in two parts: the first part will be focused on evaluating the method, while the second part will be focused on understanding the components of internship satisfaction and interns' performance.

¹⁹ As we will detail in the discussion, one of the main advantages of such a measure would be to provide information about the worker's satisfaction very early so that managers could act on it as soon as possible, solving potential problems before they become too troublesome to deal with.

Evaluating the method

Given that the main goal of the internship is to *learn*, we expect a strong correlation between feeling of learning and satisfaction. We should be able to observe it throughout the internship, hence our first hypothesis:

H1: Satisfaction is strongly correlated to learning throughout the internship.

The main objective of this new method is to be able to get a simple picture of the intern's experience as early and as simply as possible. As such, we believe that the measure of satisfaction during the first couple of months will be strongly correlated to what we measure at the end. Furthermore, if satisfaction and performance are correlated, as literature suggests, then we should also observe this link very early. This leads us to the second hypothesis, which we will decline into two sub-hypotheses:

H2A: Final satisfaction is strongly correlated to the satisfaction at the beginning of the internship.

H2B: Internship grade is strongly correlated to the satisfaction at the beginning of the internship.

Another objective of this alternative method is to highlight the shortcomings of measuring satisfaction only at the end. As we explained earlier, one of the most well-known shortcomings is the peak-end effect. This leads us to H3:

H3: Final satisfaction is strongly correlated to the satisfaction at the end of the internship and to the satisfaction of the most extreme period.

Understanding the determinants of a successful internship

Finally, the literature has shed some light on what the main determinants of satisfaction and performance are, as shown in the previous section. We expect both satisfaction and internship grade to depend on interns' expectations regarding their internship (what we call "aspirations"), as well as their detailed evaluation of the internship. Hence our two final hypotheses:

H4: Final satisfaction is predicted by evaluations and aspirations for the internship.

H5: Internship grade is predicted by evaluations and aspirations for the internship.

These broad hypotheses will give us flexibility when building models to predict those two key variables.

In the next section, we will detail how we collected the data of the study.

Part 3 – Study design

This study consisted in following a sample of one hundred and fifty-nine volunteer students during their internship. For the study, we invited students in Business, Economics, Science and Engineering who were ultimately aiming for an executive position. We chose to have a sample of interns from these fields because the wide array of activities they would perform allows for a richer study.

For this reason, we invited students from Paris 1 Panthéon-Sorbonne, EBS (European Business School), ESAA (École Supérieure Algérienne des Affaires), and AgroParisTech (a French engineering school). Thanks to word of mouth, students from other universities also offered to participate, and we welcomed them in order to have a larger sample.

Data collection

The data was collected through three different online questionnaires, sent at specific times in the study.

First, we conducted a questionnaire at the very beginning of the study, in January and February 2017²⁰. We chose January to begin the study because it is the time of the year when most students would begin their internship, right after completing the first semester's exams. We invited people to participate in our study via an e-mail which included a link to the first questionnaire. In this one, we asked for the following information: gender, age, current studies, university / school, specialization, personality (using the Ten Item Personality Inventory) (Gosling, Rentfrow, & Swann Jr, 2003), internship start and end dates, company and aspirations (via the Happy Trainees expectations questionnaire). The email invitation to participate to this study was sent before the expected start of their internship, and all those who participated did so before the beginning of their internship.²¹ The e-mail addresses were provided by the schools and universities after we explained to them the point of the study, and how it was aimed solely at improving the experience of future interns by helping us understand what makes a good internship.

Before moving on, we should clarify what we really measure with our questionnaire on "aspirations". With respect to his future internship, a student might have "wishes and expectations".

²⁰ We did not send all invitations at the same time; we contacted some schools in January and the rest in February depending on the time it took us to have access to the relevant mailing list.

²¹ We checked it by comparing the date at which they submitted an answer to the date at which their internship was to begin.

For instance, he might wish for an internship that pays well, but expect it to pay poorly and be fine with it. With this questionnaire, we measure wishes. All items begin with “I would like an internship that [...]”. Another important thing to note is the timing of that questionnaire: they answer it when they already found an internship, but it has not yet started. At this point, the student may have a good idea of what to expect, but that does not mean he revised his wishes. This is why this questionnaire should be seen as an evaluation of the values of the students, of what matters to him, and not of what he expects from the internship. In other words, answers to this questionnaire will not be influenced by the internship that the student found (the opposite is false, of course, as the student will try to find an internship that corresponds to his values).

Then, during each month of their internship, we sent them an e-mail invitation at a random time and day²². This invitation contained a link to a very short, three-question, survey²³. First, we asked them to select from a list the activities they performed during the last two days and, for each of these activities, how satisfied they were with it and how much they felt like they learned while doing it. We asked them to answer these questions as quickly as possible. This questionnaire represents the new methodology that we test in this study: a very short survey that is done *during* the moment, and not in the end of a period. This would allow us to avoid the biases presented in Kahneman *et al.* (1997), notably the peak-end effect.

Furthermore, its short nature and the fact that we ask them to answer it as quickly as they can mean that we can get an intuitive and more emotional measure of their satisfaction, as opposed to the rationalized and reflective measure that we get when we ask them to think back on their entire internship once it's done.

Finally, on the last day of their internship, we sent them a final questionnaire. This means we sent each questionnaire individually, based on the end date that the intern provided us with in the initial questionnaire. They did not know their grade yet, so they sent it to us via email at a later date. This means that it's not possible for their grade to have influenced their satisfaction. And of course, their internship tutor, the person who gives the grade, does not know the final satisfaction of the intern. This ensures that both variables can be treated as independent in our analyses. They were reminded that they would not be eligible for the financial incentive if we did not have a grade, and we left them ample time by distributing this incentive as late as possible, at the end of November 2017. In this final questionnaire, we asked them to report their overall satisfaction concerning their internship and their evaluations (using the Happy Trainees evaluation questionnaire, which mirrors

²² This random time and day were determined in advance with the help of a random number generator and were the same for every intern. They were not informed of this.

²³ It was always the exact same survey that we sent multiple times during the internship.

the aspiration questionnaire in the beginning of the survey). This allows us to know whether their initial aspirations were met, and their internship grade.

The questions that we asked them via both Happy Trainees questionnaires are available in the appendix 1, translated from the original French.

Incentives

We used three types of incentives to maintain and reinforce the student's will to carry on with the questionnaires during several months in a row: the first one was a "certificate of scientific contribution" awarded at the end of the study – a purely symbolic document. Recent literature (Kosfeld & Neckermann, 2011) has shown that this type of free incentive could be effective at improving engagement and performance, as it fulfills the basic need for recognition. The second one was a lottery in which all people who participated to the very end were enrolled, allowing them to win one of four prizes of 100 €. Lastly, we made sure they were aware that this was a scientific study whose goal was to improve the organization of future internships. This means they knew the study would serve a good cause and that it would help the future students who would later be in their place, which could activate a normative goal-oriented initiative on participants (assimilated to behaving the right way or showing exemplary behavior)²⁴.

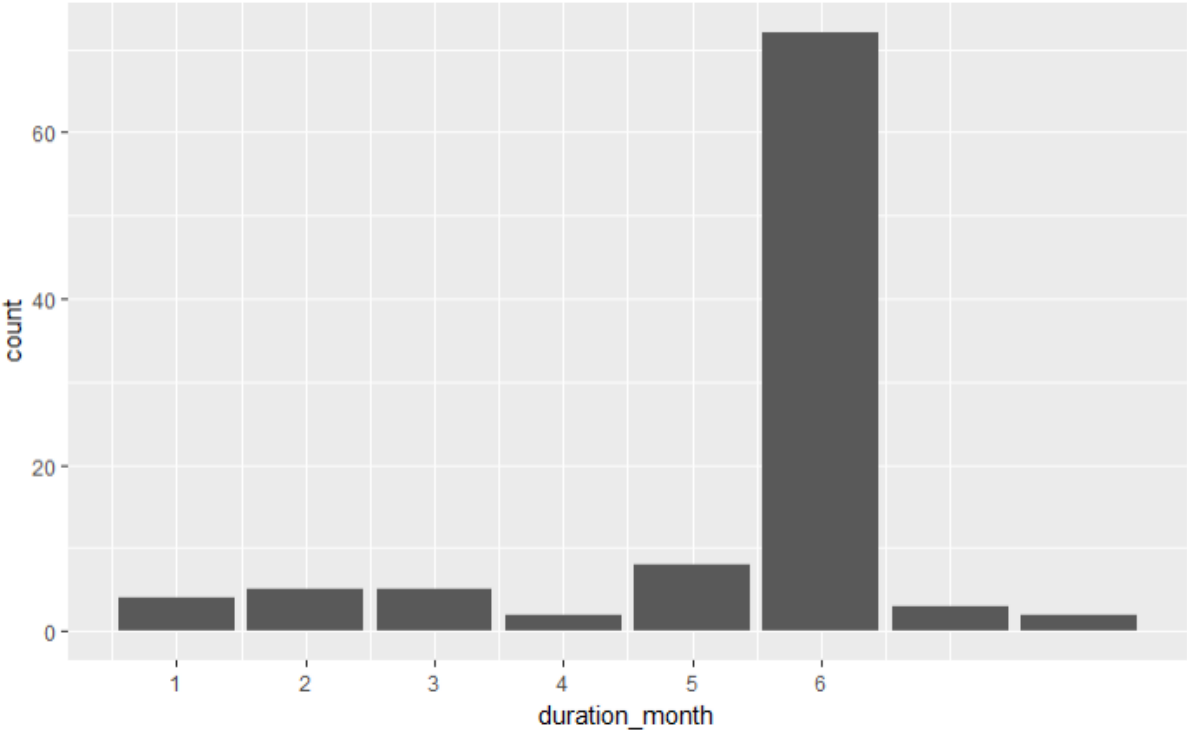
Part 4 –Results

Descriptive statistics

89% of participants are 24 years old or less, and the mean age is 23 (SD = 2). There are 48% females in the sample. Internships could last anywhere between 1 and 7 months, but 71% of the students in our sample had an internship that lasted for approximately 6 months, with a mean duration of 163 days (SD = 44 days). The distribution of durations is shown on the figure below.

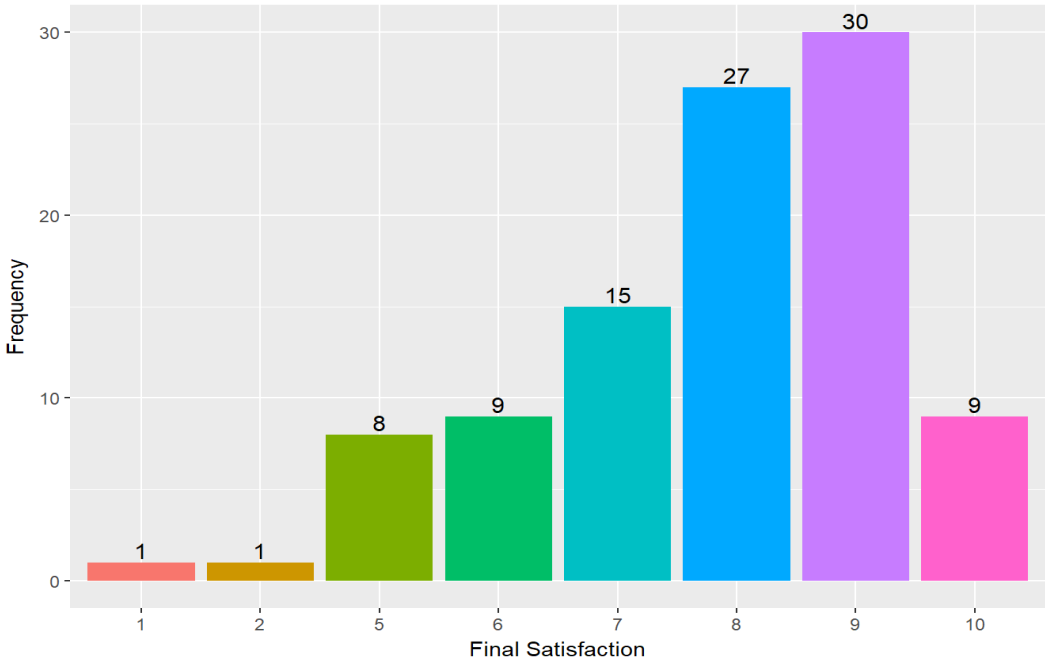
²⁴ For more, see (Cialdini, Reno, & Kallgren, 1990).

Figure 8 – Internship duration



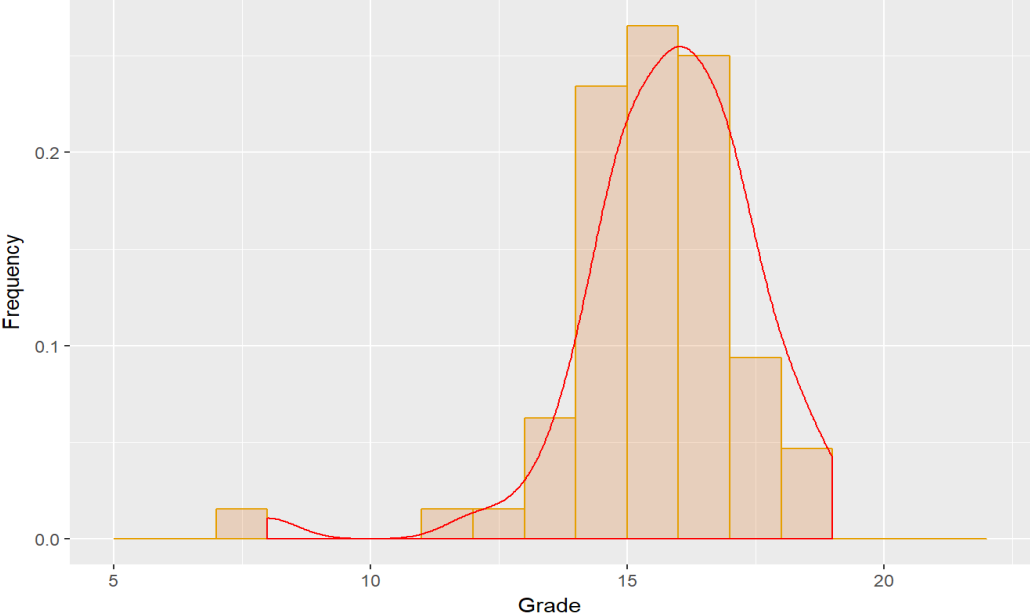
Among all participants (159 in total), only 101 fulfilled the periodical and the final questionnaires and almost all participants reported high levels of final satisfaction. As observed in the figure below, 66% rated their satisfaction at level 8 or above, with a mean final satisfaction of 7.8 (SD = 1.66).

Figure 9 – Final satisfaction



The average performance was also quite high among the 64 participants who gave us their internship grade, with most of them having a grade that is equal or greater than 15 out of 20, as shown in the figure below. The mean grade for this sample was 15.86 (SD = 1.73).

Figure 10 – Grade distribution



Shapiro-Wilk tests showed that final satisfaction and internship grade were not normally distributed, since these variables are too concentrated in the middle.

Before analyzing the results, we checked whether there was a selection bias. About 200 people were interviewed in early 2017, and the 159 people we mentioned earlier are those who had already found an internship at the time of the questionnaire. It is thus entirely possible that these students significantly differed from the others.

To do that, we performed a Probit regression in which the dependent variable was whether or not the respondent participated through the end of the study (as opposed to stopping anywhere before that point). This analysis revealed that only people who saw themselves as “disorganized/neglectful” were significantly ($p < 0.05$) less likely to participate until the end of the study, a variable which isn’t significant in any of the other regressions, so we consider that selection bias would not be an issue for this study.

Evaluating the method

Satisfaction and learning

We started by observing whether, in each month, there was a correlation between how satisfied people were with the activities they just performed, and how much they thought they learned while doing them. We noticed that the feeling of learning (that we will simply refer to as “learning” from now on) was very strongly correlated to the satisfaction. This correlation remained steady throughout all six months, as shown in the table below. We expected interns to care deeply about how much they learned since this is the very purpose of an internship, but the magnitude and consistency of the correlation remains remarkable. This really shows that companies should make sure that the interns have as many opportunities to learn as possible.

Table 18 - Monthly correlation between reported learning and monthly internship satisfaction

| | | | | | |
|----|----------------|----|----------------|----|----------------|
| ## | March | ## | April | ## | May |
| ## | pearson 0.6 | ## | pearson 0.5992 | ## | pearson 0.645 |
| ## | pvalue 0.0 | ## | pvalue 0.0000 | ## | pvalue 0.000 |
| | N = 62 | | N = 71 | | N = 81 |
| ## | June | ## | July | ## | August |
| ## | pearson 0.6205 | ## | pearson 0.6397 | ## | pearson 0.6739 |
| ## | pvalue 0.0000 | ## | pvalue 0.0000 | ## | pvalue 0.0000 |
| | N = 56 | | N = 43 | | N = 39 |

This statement is in complete agreement with most literature assertions (Hong, 2002), (Lee, Srinivasan, Trail, Lewis, & Lopez, 2011), (Johnson, Aragon, Shaik, & Palma-Rivas, 2000).

Most of the internships lasted for 6 months, but not all. As such, June could be the first month for one subject and the last month for another. This means we cannot interpret the satisfaction and learning scores further without first transforming the data. We thus created another series of variables to observe the evolution of satisfaction and learning over time.

The variables sat_1 to sat_6 refer to the average satisfaction reported by an intern from the first to the 6th questionnaire he answered. Learn_1 to learn_6 refer to the average learning reported in these questionnaires. Since interns did not all answer the same number of questionnaires (a few of them had an internship that lasted less than 6 months and thus were interrogated less than 6 times, while others ignored the questionnaires after some time or only answered sporadically), it means

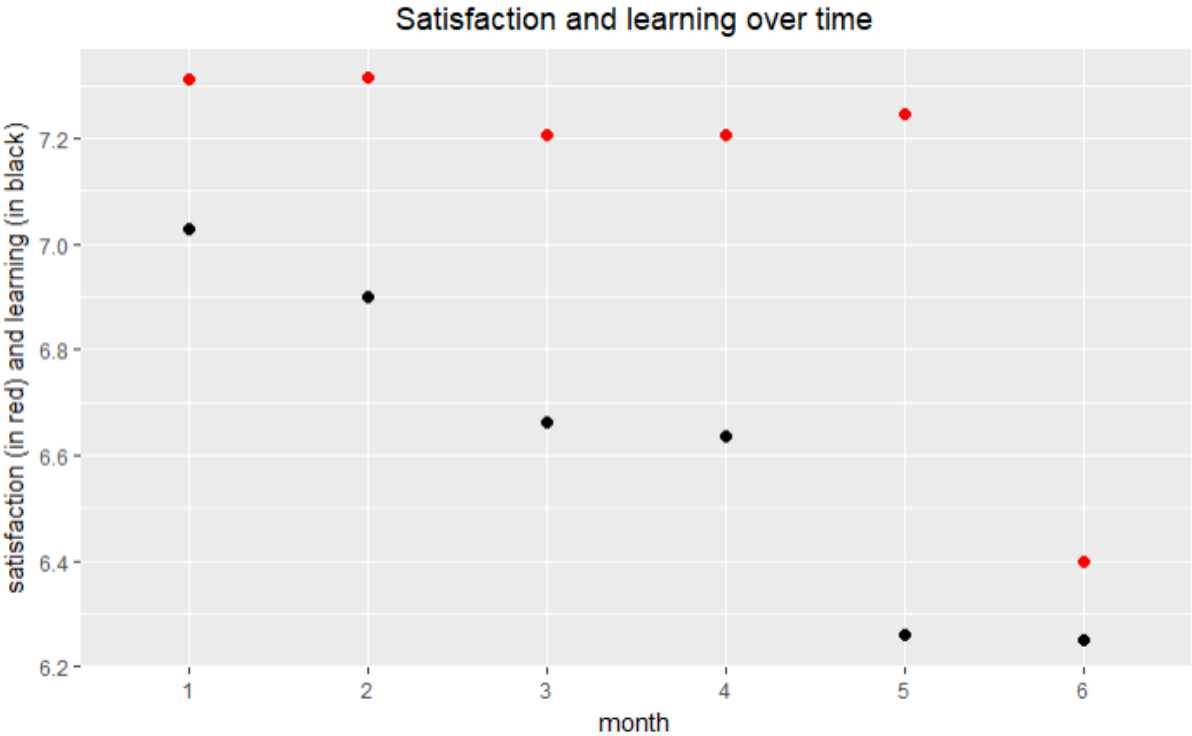
that the sample size gets smaller with each increment. The table below shows the sample size that we have for each number of answers:

Table 19 – Sample size for each number of answers

| | | | | | | |
|--------------------|----|----|----|----|----|---|
| Answered at least: | 1 | 2 | 3 | 4 | 5 | 6 |
| Sample size | 87 | 77 | 64 | 49 | 30 | 9 |

This means that while 87 people answered at least once, only 9 answered all 6 times. The graph below shows the evolution of these two variables over time, where the average satisfaction is in red and the average learning is in black, while “month” indicates which questionnaire this answer is from in chronological order:

Figure 11 – Satisfaction and learning over time



We can see that, with the exception of the last period, satisfaction remains steady over time while learning decreases, which is consistent with what we could reasonably expect: as time goes by, the interns will have less to learn. The surprising results of period 6 are simply due to the very low number of people who did answer 6 times. Indeed, answering all 6 times require a strong sense of discipline and consistency since it means that the subject missed absolutely no monthly

questionnaire. Unfortunately, because of this low sample of people who answered 6 times, we cannot make any meaningful comparison between them and the others.

Nevertheless, this decrease in satisfaction remains small (approximately a 1-point difference in satisfaction), and is negligible if we discard sat_6. It would still seem that learning and satisfaction are highly correlated. The Pearson’s correlations for each period are displayed in the table below:

Table 20 – Correlation between satisfaction and learning over time

| | | | | | | |
|-------------|------|------|------|------|------|------|
| Period | 1 | 2 | 3 | 4 | 5 | 6 |
| Correlation | 0.66 | 0.62 | 0.56 | 0.63 | 0.56 | 0.38 |

This shows that learning is indeed highly correlated to satisfaction. This result is consistent with the literature (Daugherty, DeWitt C. Baldwin, & Rowley, 1998), and the relationship between these two variables may be amplified by the fact that the sole purpose of an internship is to learn new skills, as we said earlier. This table also verifies the fact that the correlation between these two variables decreases over time.

In the end, we can accept H1: there is indeed a very strong correlation between satisfaction and learning throughout the internship.

Preparing the data for further testing

The previous data transformation allowed us to observe evolutions over time, but we need to go a step further in order to test the next hypotheses. Since the sample of people who answered is different for each month, the scope of the analyses we can perform with our current set of variables is limited.

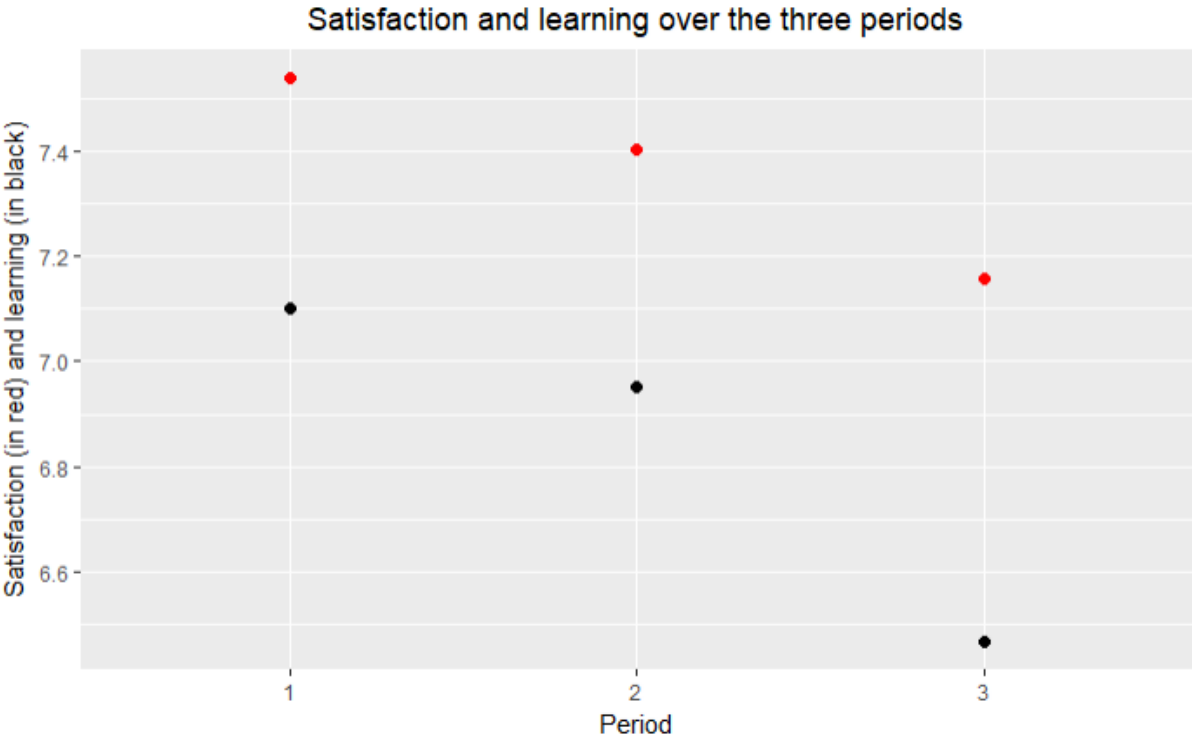
This is why we selected a subsample of people who answered at least once every two months, and constructed three measures of satisfaction for them, one for each couple of months over the duration of their internship. We used the following method: if the subject answered both months, we computed the mean of both satisfactions, and if he only answered one out of two months, then we used that answer as a measure of satisfaction.

For example, the internship of one person lasted from April to September, and he answered in April, May, June and September. To compute the first value, that we call sat_beginning, we calculated the mean of his answers in April and May, which are the beginning of his internship. To compute the second value, sat_middle, we took the answer of June (because he did not answer in

July), which is in the middle of his internship, and to compute the third value, sat_end, we took the answer of September (because he did not answer in August), which is the end of his internship.

We chose this method because the majority of people in our sample was interviewed during 6 months, and none were interviewed more than 6 months. This is the method that allowed us to take a sample as large as possible (48 people) while having no missing values. In the figure below, we can see how satisfaction and learning (which was computed in the exact same way) evolve over time.

Figure 12 – Evolution of satisfaction and learning over the three periods of the internship



We can see that the trend remains largely similar as the one that was observed before, in that satisfaction and learning decrease over time. However, the decrease observed here is small. In the next subsection, we will now use these new variables to investigate the existence of a peak-end effect.

Peak-end effect, and the impact of the different periods on the outcomes of the internship

In order to investigate the peak-end effect, we chose to build a table of correlations between each of the three measures of monthly satisfaction, the peak satisfaction, final satisfaction and

internship grade. The peak satisfaction is the maximum satisfaction score that the intern has ever expressed in a monthly survey²⁵. The results are shown below:

Table 21 – Correlations between different measures of satisfaction and internship grade

| | sat_beginning | sat_middle | sat_end | final_satisfaction | Peak_sat |
|--------------------|---------------|------------|---------|--------------------|----------|
| sat_beginning | | | | | |
| sat_middle | 0.45** | | | | |
| sat_end | 0.36* | 0.60*** | | | |
| final_satisfaction | 0.60*** | 0.27 | 0.37* | | |
| Peak_sat | 0.69**** | 0.54*** | 0.60*** | 0.57*** | |
| internship_grade | 0.07 | 0.04 | 0.07 | 0.42* | -0.01 |

p < .0001 '*****'; p < .001 '****', p < .01 '***', p < .05 '**'

Here we can see that final satisfaction is mainly driven by the satisfaction reported at the first period. It is also significantly predicted by satisfaction at the last period, but the correlation is much weaker. We also see that the effect of the peak remains the same: there is a significant correlation between the peak satisfaction and the final satisfaction.

This is quite interesting because, while it verifies our intuition, it goes beyond what the peak-end effect usually shows (which is that the satisfaction at the last period has a larger impact on the final satisfaction than *any* other period's). This means that in the case of these internships, the very beginning is of utmost importance, and companies should really ensure that interns can learn a lot from the start. The peak effect, however, is as usual, with the highest satisfaction being highly correlated to the final satisfaction. Finally, we can see that none of the monthly satisfaction measures is correlated to the internship grade.

In light of these results, we can accept H2A. Final satisfaction is indeed strongly correlated to the satisfaction at the beginning of the internship. However, we must reject H2B. There is no correlation between the internship grade and the satisfaction at the first period. These results also verify the peak-end effect, so we can accept H3. Final satisfaction is indeed correlated to the last and peak satisfactions.

²⁵ Of course, one could consider that using the maximum is hardly a good measure of the intensity of an experience. However, subject retention is a major issue in a study as long and demanding as this one, and we did not want to risk losing more participants by adding another question. Besides, as imperfect as it is, the maximum is the standard way of evaluating intensity in studies about the peak-end effect, such as the original one (Kahneman et al., 1997).

In order to further validate the effect of different periods on final satisfaction, we conducted a series of correlation tests between the satisfaction for each period of the internship and the final satisfaction. In the table below, sat_monthn is the satisfaction score given by people during the nth month of their internship, or the nth time they were interviewed²⁶. Note that, each month, a different number of people interviewed. Furthermore, as time went on, less and less people were interrogated (as not everybody had an internship that lasted 6 months), so the sample size decreases. However, among people who *were* interrogated, the participation rate remains relatively steady over time.

Table 22 – Correlations between monthly satisfactions and final satisfaction

| | Correlation with final satisfaction | Sample size |
|------------|-------------------------------------|-------------|
| Sat_month1 | 0.47*** | 69 |
| Sat_month2 | 0.33** | 66 |
| Sat_month3 | 0.24 | 63 |
| Sat_month4 | 0.2 | 49 |
| Sat_month5 | 0.44** | 40 |
| Sat_month6 | 0.53*** | 33 |

p < .0001 '*****'; p < .001 '***', p < .01 '**', p < .05 '*'

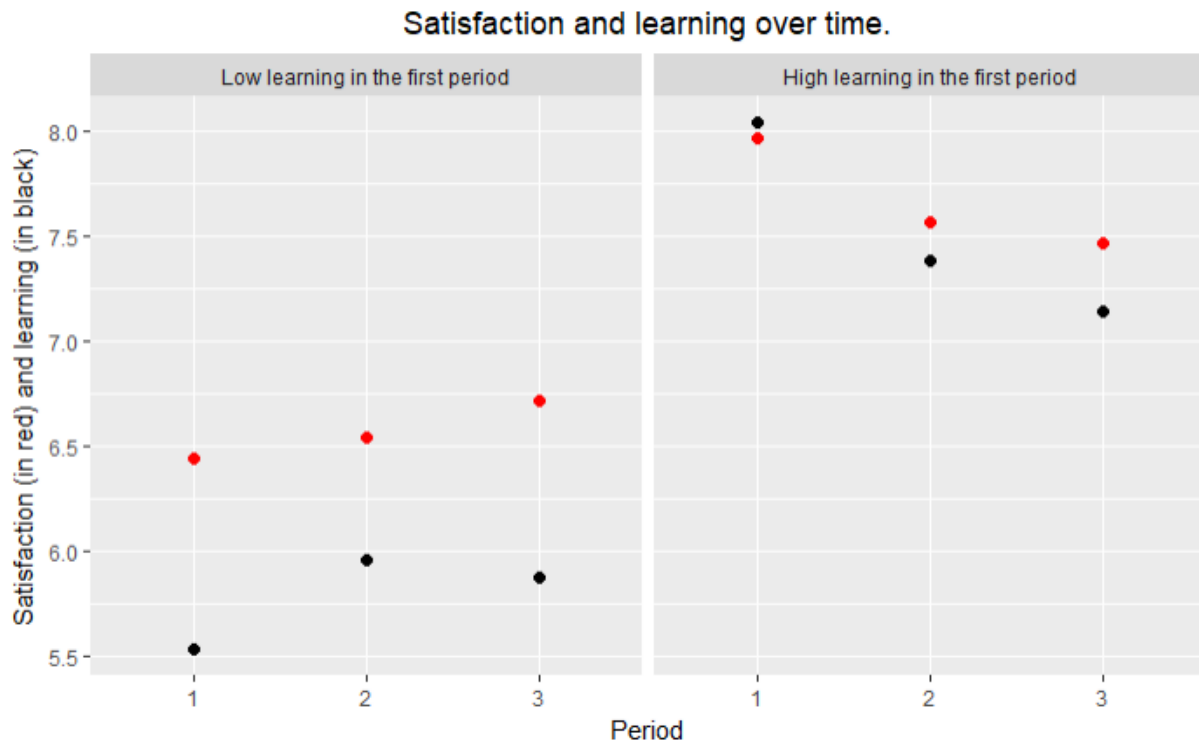
Here, we see that, contrary to what we observed when performing the transformation of satisfaction variables, the ending and the beginning of the internship both have the same effect on final satisfaction, with months in the middle having no effect on it. It thus indicates that final satisfaction is really driven by the last and first memories of the internship, which is evidence of what we might call a “beginning-peak-end” effect.

Do people who learn a lot in the first period have a different profile?

In order to complement what our hypothesis testing taught us, we did a subgroup analysis with the three period variables (sat_beginning, sat_middle and sat_end) in order to see if different groups of people had different trends. First, we compared people who reported learning a lot during the first period to those who reported not learning much. The median of the learn_1 variable was 7, so this is the value we used to split the two groups, those answering 7 or more being represented on the plot on the right:

²⁶ With the previous measure, sat_n, we recorded the scores the nth time they were *answered*. With the current measure, sat_monthn, we recorded the scores the nth they were *interrogated*. This means that, for example, the sample for sat_5 is the sample of people who answered 5 times, while sat_month5 is the sample who answered the 5th monthly questionnaire we sent them.

Figure 13 – Satisfaction and learning over time based on initial learning. Satisfaction in red.



On this graph, it would seem that satisfaction increases slightly over time for people who reported not learning a lot in the first period, while it decreases steadily over time for people who reported learning a lot in the first period. To verify this impression, we conducted Student tests to compare the first monthly satisfaction to the last monthly satisfaction for each of the two groups.

We find that, for the first group, there is no difference between the first monthly satisfaction ($M = 6.44$, $SD = 1.17$) and the last monthly satisfaction ($M = 6.71$, $SD = 1.06$, $p\text{-value} = 0.3282$). However, for the second group, we find that the first monthly satisfaction ($M = 7.97$, $SD = 1.1$) is significantly higher than the last monthly satisfaction (7.46 , $SD = 1.04$, $p\text{-value} = 0.02853$).

The results are similar if we check the difference between the feeling of learning during the first period to the feeling of learning in the last period, which is consistent with the fact that the feeling of learning and satisfaction are strongly correlated.

This means that people who reported learning more at the beginning of the internship tended to become less satisfied about it as time went by. This is likely because they learned so much at the beginning that things could only go down as time went by. And since the later months couldn't match the sheer amount of knowledge and skills they acquired at the beginning, they naturally became a bit less satisfied. On the other hand, those who have a more balanced learning experience throughout the internship will also have a more stable satisfaction, since satisfaction and learning are so strongly

correlated. However, this satisfaction will be lower overall. It is important to note that even their last monthly satisfaction score is still significantly higher than the last monthly satisfaction score of people who reported learning less at the beginning of their internship (Student test, p-value = 0.003129). In the end, our recommendation for companies which hire interns would be to try to make sure they can learn as much as possible from the beginning, since the first few months are extremely important. However, they should also make sure to keep teaching interns, to avoid the decrease in satisfaction that we observed. This would ensure a more stable satisfaction which might leave them with a fonder memory of the company overall.

This result is very strong because it means that, combined with the fact that the satisfaction at the first period has the highest correlation to final satisfaction, we can basically know right from the beginning how well the internship will go. It is not necessary to wait until the assessment at the end. In other words, if such a tool was at the disposal of human resources, they could then act immediately to correct the trajectory of the internship. It also goes to show that a lengthy questionnaire at the end is unnecessary, as these very simple measures are just as good at assessing the quality of the experience.

Finally, there is no difference in average internship grade between the two groups, which means that this difference in trend did not appear to have affected their performance. We also checked for differences in aspiration levels between the two groups, but Student tests revealed nothing significant.

The role of the activities performed

As a reminder, in the monthly questionnaires, subjects were asked not to evaluate their last 2 days as a whole, but to evaluate each activity of the last two days based on how satisfied they were and how much they learnt. This component of the study was purely exploratory, which is why we did not formulate any hypothesis regarding it. This allows us to build the following table, which shows how much satisfaction and learning varies based on the particular activity they performed.

Table 23 – Satisfaction and learning per activity

| Activity | Number of answers | Satisfaction | Learning |
|--|-------------------|--------------|----------|
| Data collection and analysis | 300 | 7.36 | 6.9 |
| Working on new projects and innovations | 135 | 7.73 | 7.62 |
| Customer relationship management | 68 | 7.29 | 7.41 |
| Auditing | 10 | 7.7 | 7.7 |
| Purchase management | 13 | 7.23 | 7.54 |
| Working on marketing projects and sales | 21 | 7.65 | 7.27 |
| Human resources management | 13 | 6.92 | 7.08 |
| Accounting and finances | 24 | 6.67 | 6.88 |
| Administrative work | 101 | 5.48 | 3.93 |
| Company event | 40 | 7.88 | 6.45 |
| Attending formations | 23 | 8.09 | 8 |
| Other | 75 | 7.81 | 7.29 |

Once again, we find in this table one of our core results: satisfaction is heavily correlated to learning. It is interesting to note that, with one exception, satisfaction and learning do not vary significantly from task to task. The single exception is administrative work, which is rated significantly lower in satisfaction and even more so in learning according to a Student test of differences. It is interesting to note that the loss in satisfaction does not have the same magnitude as the loss in learning. This could mean that learning is a less biased indicator.

It is also interesting to note that only administrative work differs from the rest of the table. It seems obvious that it would be rated lower, but one might not expect such a low variance among the rest of the activities. This suggests that what exactly they do isn't as important as simply giving interns the opportunity to learn marketable skills.

Understanding the determinants of a successful internship

Now that we have presented the results of our method, we will move on to what the rest of the study can teach us about the components of a successful internship.

Before doing any further work, we wanted to investigate answers to the Happy Trainees questionnaires. More specifically, we wanted to know whether there was any correlation between the different measures of aspirations and evaluations, in addition to how much each of them was correlated to our main variables of interest, which are final satisfaction and the internship grade. It is important to know about this now as strong correlations might translate to multicollinearity problems in future regressions. The table below shows the correlations between these variables²⁷:

Table 24 - Correlation between evaluations, final satisfaction and internship grade

| | eval_envir | eval_pride | eval_motivation | eval_fun | eval_management | eval_progres | final_satisfaction |
|--------------------|------------|------------|-----------------|----------|-----------------|--------------|--------------------|
| eval_envir | | | | | | | |
| eval_pride | 0.60**** | | | | | | |
| eval_motivation | 0.48**** | 0.53**** | | | | | |
| eval_fun | 0.59**** | 0.68**** | 0.59**** | | | | |
| eval_management | 0.54**** | 0.60**** | 0.66**** | 0.67**** | | | |
| eval_progres | 0.25* | 0.60**** | 0.54**** | 0.71**** | 0.64**** | | |
| final_satisfaction | 0.43*** | 0.60**** | 0.66**** | 0.74**** | 0.65**** | 0.67**** | |
| internship_grade | 0.01 | 0.25* | 0.27* | 0.13 | 0.23 | 0.14 | 0.29* |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

In Happy Trainees questionnaires, there are 6 groups of 3 items, each group measuring one dimension among the following: motivation, environment, pride, fun, progression and management. The items in the questionnaire on evaluation mirror the items in the questionnaire on aspirations.

The most striking result is that all evaluations measures are highly correlated to one another and to final satisfaction. It appears that interns evaluate their internship in a more holistic way in the end. If they are satisfied overall, then they are satisfied about every aspect, and vice versa. This also means that we should be wary of including multiple evaluation measures in regressions. While doing so is technically possible, it would not allow us to provide an accurate interpretation of the coefficients.

Two dimensions stand out from the rest because of their correlation to the internship grade: “pride” and “motivation”. Interestingly, there is no significant correlation between how much one felt like they learned during the internship and their performance. We will explore the implications of this result in the discussion. Note also that the correlations here are much lower than between satisfaction and evaluations.

²⁷ Note that, in order to include “internship grade” in this table, we had to limit ourselves to the sample of individuals who gave us an internship grade (64 people). We also run the correlations between evaluations, aspirations and final satisfaction on the sample of people who participated to at least the last questionnaire (100 people), and the results are identical.

In the table below, we now observe the correlation between aspirations, final satisfaction and the final grade.

Table 25 - Correlation between aspirations, final satisfaction and internship grade C

| | aspi_envir | aspi_pride | aspi_motivation | aspi_fun | aspi_management | aspi_progres | final_satisfaction |
|--------------------|------------|------------|-----------------|----------|-----------------|--------------|--------------------|
| aspi_envir | | | | | | | |
| aspi_pride | 0.30* | | | | | | |
| aspi_motivation | 0.16 | 0.37** | | | | | |
| aspi_fun | 0.24 | 0.39** | 0.44*** | | | | |
| aspi_management | 0.48**** | 0.53**** | 0.41*** | 0.26* | | | |
| aspi_progres | 0.02 | 0.30* | 0.36** | 0.40** | 0.26* | | |
| final_satisfaction | -0.04 | 0.04 | 0.22 | 0.20 | 0.07 | 0.17 | |
| internship_grade | 0.18 | 0.06 | 0.13 | -0.05 | 0.27* | -0.14 | 0.29* |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Correlations between aspiration domains are still highly significant overall, but not as high as correlations between evaluations. However, with the exception of “motivation”, there is no correlation between aspirations and final satisfaction. We could expect people with higher aspirations to have a lower final satisfaction, as it would give them more margin for disappointment, but this is not the case, probably because they also had higher standards when looking for an internship.

Only the aspiration for “management” is significantly correlated to the final grade. This group is about the desire for recognition as well as having clear objectives and an “inspiring manager”. In order to further investigate this result, we checked for correlation between each of the individual items and the internship grade, but none of these correlations is significant, and items from this group are correlated to each other, which means it is the measure for the group as a whole that matters. Our interpretation of this is that people who desire recognition and clear objectives, among other things, will be people who typically want to work hard in order to impress their boss – especially if they have one that they find “inspiring”.

Next, we investigated whether all individual items did represent a unique construct. In order to make sure of this, we ran a series of correlation tests between the items in each group. It is important to investigate this prior to using the variables in a regression because this allows us to know if each group does measure a single dimension (in which case all items are correlated), or if a single group actually measures different dimensions (in which case we would observe little to no correlation between the items). In the second case, it would thus make sense to include the individual items rather than the group in a regression.

The detailed results are in the appendix. In general, we observe a moderate to strong correlation between the different evaluation items, a weak correlation between the aspiration items,

and no correlation between aspirations and evaluations. However, there are a few exceptions. We never observe a negative correlation. This shows that groups of items are not all that homogeneous and, in the rest of the analysis, we will remember these results when deciding whether the group or the items should be included in the analysis.

Now that we have analyzed the aspirations and evaluations, we can start building models to predict final satisfaction and the internship grade.

Predicting final satisfaction

As explained before, to avoid issues with multicollinearity, we will only include one measure of aspiration and one measure of evaluation in each model, so it is just a matter of picking which one to keep. Regarding final satisfaction, we assumed that it would mainly be influenced by how much one felt like they learned during the internship, since learning is the main purpose of the internship. Final satisfaction should also be influenced by the aspirations that one has for the internship. We chose to keep the aspiration for “fun” because it is about the desire for a project to which they can be 100% committed and the desire to have an activity in relation to one’s interests and values. Such aspirations suggest that the intern actually enjoys working in his field and will look for an internship that he might genuinely enjoy, and so it is likely that it would positively influence his satisfaction. We then added to the model personality, university, age and gender as control variables.

The resulting model is shown in the table below:

Table 26 – First regression of final satisfaction

| final satisfaction | | | |
|--|------------------|--------------|------------|
| <i>Predictors</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 0.86 | -3.26 – 4.98 | 0.682 |
| age | -0.03 | -0.16 – 0.10 | 0.626 |
| female | 0.07 | -0.51 – 0.65 | 0.823 |
| Agro Paris Tech | -0.17 | -1.08 – 0.75 | 0.724 |
| Sorbonne | -0.59 | -1.72 – 0.55 | 0.314 |
| ESAA | -0.87 | -1.91 – 0.16 | 0.103 |
| Other univ | -0.10 | -1.14 – 0.94 | 0.847 |
| extraversion | 0.09 | -0.01 – 0.19 | 0.075 * |
| agreeableness | -0.09 | -0.22 – 0.05 | 0.209 |
| consciousness | 0.07 | -0.04 – 0.19 | 0.204 |
| emotional stability | 0.05 | -0.04 – 0.15 | 0.272 |
| openness | 0.05 | -0.05 – 0.16 | 0.330 |
| eval progres | 1.62 | 1.19 – 2.05 | <0.001 *** |
| aspi fun | -0.39 | -0.84 – 0.06 | 0.093 * |
| Observations | 100 | | |
| R ² / adjusted R ² | 0.545 / 0.477 | | |

Here we can see that the only explanatory variable that is significant at the 5% level is how much one felt like they learned during their internship, and that its impact is fairly large, as could be expected from the correlations we observed in the previous section.

Then, we re-run the previous model after removing every variable that was not significant at least at the 10% level. Our objective is to arrive at the simplest model possible for explaining our key variables, which is why we chose to only keep those who have a significant effect. The new model is shown in the table below:

Table 27 - Concise model of final satisfaction

| final satisfaction | | | |
|--|------------------|---------------|------------------|
| <i>Predictors</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 1.00 | -0.88 – 2.88 | 0.302 |
| extraversion | 0.09 | 0.00 – 0.18 | 0.044 |
| eval progres | 1.73 | 1.36 – 2.10 | <0.001 |
| aspi fun | -0.48 | -0.87 – -0.10 | 0.015 |
| Observations | 101 | | |
| R ² / adjusted R ² | 0.504 / 0.489 | | |

The first thing we can see here is that having a high aspiration for “fun” actually *decreases* final satisfaction, which goes contrary to our assumptions and to what the correlation tables of the previous sections have shown us. We can also see that the Adjusted R-Squared is approximately the same despite a much lower number of variables.

In order to better understand the result regarding the effect of the aspiration for fun, we run three additional regressions. In each one, we replaced the variable “aspi fun” by the score for one of the three items that make up the group. The results are presented in the table below:

Table 28 – Final satisfaction predicted by three different aspiration measures of “fun”

| <i>Predictors</i> | final satisfaction | | | final satisfaction | | | final satisfaction | | |
|--|---------------------------|--------------|------------------|---------------------------|--------------|------------------|---------------------------|---------------|------------------|
| | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | -0.48 | -2.40 – 1.44 | 0.623 | 0.24 | -1.48 – 1.97 | 0.783 | 0.40 | -1.26 – 2.07 | 0.636 |
| extraversion | 0.09 | 0.01 – 0.18 | 0.040 | 0.10 | 0.01 – 0.19 | 0.026 | 0.07 | -0.01 – 0.16 | 0.096 |
| eval progres | 1.61 | 1.23 – 1.98 | <0.001 | 1.70 | 1.32 – 2.08 | <0.001 | 1.68 | 1.32 – 2.04 | <0.001 |
| aspi fun 1 | 0.10 | -0.15 – 0.35 | 0.444 | | | | | | |
| aspi fun 2 | | | | -0.22 | -0.46 – 0.02 | 0.070 | | | |
| aspi fun 3 | | | | | | | -0.31 | -0.51 – -0.10 | 0.004 |
| Observations | 101 | | | 101 | | | 101 | | |
| R ² / adjusted R ² | 0.476 / 0.460 | | | 0.491 / 0.475 | | | 0.517 / 0.502 | | |

aspi fun 1: “I want to perform an activity in relation to my interests and values”

aspi fun 2: “I want to be 100% committed to a project”

aspi fun 3: “I want to join a real community of interns”

We find that item 1, is not significant. Item 2 is only significant at the 10% level, and has a weak effect. It is item 3 which has a large and highly significant effect on final satisfaction. This

means that it is that last item that drives the result we observed above. The most obvious explanation for this last effect is that students who expect to find a strong community of interns when they arrive have a lot of room for disappointment, but another possibility is that this item simply captures a reference effect. In other words, the questionnaire did not measure the aspiration level that really mattered, and so this effect has been captured by this item instead. Nevertheless, this shows that aspirations do matter. Since it is mainly this last item that is responsible for the effect on final satisfaction, it is what we will keep in our final model.

Lastly, we investigated the effect of “eval progress” in the same manner. The results are in the appendix. Unlike what we observed for the aspiration, all “progress” items significantly predict final satisfaction, and the Adjusted-R² of each model is much lower than the Adjusted-R² we had when we kept these items as a group. Thus, “eval progress” works better as a group in a regression.

This leads us to our final model for predicting final satisfaction using only items from the first and last survey on our study:

Table 29 – Final model of final satisfaction predicted by personality, evaluation and aspiration

| final satisfaction | | | |
|--|------------------|---------------|----------|
| <i>Predictors</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 0.40 | -1.26 – 2.07 | 0.636 |
| extraversion | 0.07 | -0.01 – 0.16 | 0.096 |
| eval progres | 1.68 | 1.32 – 2.04 | <0.001 |
| aspi fun 3 | -0.31 | -0.51 – -0.10 | 0.004 |
| Observations | 101 | | |
| R ² / adjusted R ² | 0.517 / 0.502 | | |

One might notice the slight change in the coefficient for “extraversion”. It is simply due to the obvious fact that people who want to be part of a community of interns will tend to be more extraverted. In light of this last table, we can accept H4: final satisfaction can be predicted best by the overall feeling of learning and by one measure of aspiration.

Now, we will try to predict internship grade.

Predicting the internship grade

As explained earlier, we assume that performance can be represented by the internship grade. The grade we collected was the one given by the supervisor at the end of the internship. It is not without issues of course, as supervisors will tend to give high grade to interns no matter what as a way of thanking them for their contributions. Nevertheless, it is still a numerical evaluation of performance, and there is a high enough variance between the subjects for us to use it in a regression. Note that, in this subsection and in every tests or models in which internship grade is a variable, we perform the analyses on the sample of 64 interns who gave us their internship grade.

Among the variables that we measured in the first and last survey, we assume that aspirations and evaluations should be the main determinants of the grade. Some aspirations may be important because they indicate that the intern will put more efforts into the work than others. For instance, someone who cares deeply about being 100% committed to a mission (item 2 of the “fun” group) should work harder than someone who doesn’t. As for evaluations, they represent how well certain aspects of the internship went from the point of view of the subject, so while an evaluation itself cannot impact the grade, the underlying variable that it measures can. For example, we can assume that someone who says they learned a lot during the internship would be someone who has done a lot of challenging work, and thus would have brought more to the company. Besides, one needs to be invested into one’s work in order to really learn a lot from it. This is why we want to check the impact of both evaluations and aspirations on the internship grade. We will also add personality, age, gender and university as control variables.

Of course, as we explained before, high correlation between aspirations and evaluations means we should avoid putting different measures in the same regression. We chose to add the aspiration for “management” because of what we explained when studying correlations between aspirations and internship grade: people who desire recognition and clear objectives, among other things, will be people who typically will want to work hard in order to impress their boss – especially if they care about having one they find “inspiring”. As for evaluation, we chose to test the evaluation for “motivation” because the items are most representative of a highly driven individual (with items such as “I felt committed to my mission” for instance).

After adding the same control variables that we used for the first regression of final satisfaction, we get the following model:

Table 30 – First regression of internship grade

| internship grade | | | |
|--|------------------|--------------|------------|
| <i>Predictors</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 18.03 | 8.71 – 27.35 | <0.001 *** |
| aspi management | 0.56 | -0.09 – 1.22 | 0.098 * |
| eval motivation | 0.68 | -0.04 – 1.41 | 0.071 * |
| age | -0.23 | -0.65 – 0.19 | 0.283 |
| female | 0.03 | -1.16 – 1.22 | 0.961 |
| Agro Paris Tech | 0.32 | -1.62 – 2.26 | 0.749 |
| Sorbonne | 1.39 | -0.98 – 3.77 | 0.256 |
| ESAA | 1.19 | -1.23 – 3.62 | 0.340 |
| Other univ | 0.64 | -1.47 – 2.74 | 0.557 |
| extraversion | 0.03 | -0.17 – 0.22 | 0.804 |
| agreeableness | -0.07 | -0.30 – 0.16 | 0.569 |
| consciousness | 0.09 | -0.16 – 0.34 | 0.484 |
| emotional stability | -0.07 | -0.25 – 0.11 | 0.459 |
| openness | -0.15 | -0.34 – 0.04 | 0.124 |
| Observations | 63 | | |
| R ² / adjusted R ² | 0.266 / 0.071 | | |

Here we can see that only the aspiration and the evaluation measures we chose are significant at the 10% level. None of the other variable has any significant impact. For the sake of simplicity and because none of the other variables have any theoretical importance, we chose to only keep aspiration and evaluation in future models.

We then performed the same analysis as in the previous section. We first computed a model that only included the measures of aspiration and evaluation from the model above, and then we decomposed each of these two measures into its three items. We found that “aspi management” form a homogeneous construct, so we kept the variable as it is. However, we found that the effect of “eval motivation” was entirely driven by the first item of this dimension, which was “I felt committed to my mission”. As a result, we only kept this item in the final model. The details of this analysis are in the appendix. The final model is shown in the table below:

Table 31 – Final model for predicting internship grade

| <i>Predictors</i> | internship grade | | |
|--|-------------------------|--------------|--------------|
| | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 11.67 | 8.90 – 14.44 | <0.001 |
| aspi management | 0.52 | 0.01 – 1.03 | 0.049 |
| eval motivation 1 | 0.54 | 0.02 – 1.06 | 0.048 |
| Observations | 64 | | |
| R ² / adjusted R ² | 0.130 / 0.102 | | |

In the end, the performance, just like final satisfaction, is significantly driven by aspirations. However, it is still harder to predict based on the variables we have in the first and last questionnaire than final satisfaction, as illustrated by the low Adjusted R².

In light of these results, we accept hypothesis H5: internship grade is indeed predicted by a measure of evaluation and by a measure of aspirations. In the next section of this paper, we will now discuss our findings.

Part 5 - Discussion

The results of this study have exceeded our expectations. What we measured in the first period was actually a very good indicator of how well the rest of the internship went. It means that from the first month, we can know how much an intern will learn and how satisfied he will be about the internship as a whole. And while monthly reports of satisfaction were not directly correlated to the internship grade, they could still be used to predict performance since they were highly correlated to final satisfaction, and final satisfaction was correlated to performance.

In this study, information about the monthly surveys was not sent to the company. But what if it was? If we detected early that things were not ideal, and this was communicated to the Human Resources, then perhaps measures could have been taken and the rest of the internship would have been drastically different.

We believe that this alone shows the importance of such an early and instantaneous measurement of satisfaction. It is simple and less heavy than the lengthy final questionnaire, and yet it gives most of the information needed very early, when it's still time to do something about it.

Of course, while we performed this study on interns, our ultimate goal is to generalize our findings to other persons in the workplace. The most direct extension is people who are hired in their first job: just like interns, we can assume that they are driven by a desire to learn and acquire skills that will lead to a fulfilling career. And because of their lack of professional experience, the first few months of work will be of utmost importance since it will be when they would learn the most. And if they do not learn enough during the beginning, it will be more difficult to make up for it later on. That is why we believe that a method such as ours would be as beneficial for them as it would be for interns. And obviously it would be beneficial to the employer as well, who would know that something is wrong very early and would have the opportunity to correct it immediately, before the problem becomes too well ingrained.

Interestingly, our results also are partly in contradiction with the peak-end effect. The end matters more than the middle, indeed, but it does not matter more than the beginning. This means that it isn't a universal effect when measuring satisfaction, but rather a situational one. This implies that comparing both measures of satisfaction, like we did, but in different contexts, would be a useful endeavor and a good opportunity for future research. It would allow us to shed more light on the phenomenon by highlighting where it works and where it does not, or works differently.

We believe that there are two main limitations to this study: its low sample size and its somewhat unreliable measure of performance. Finding and keeping participants for this study proved to be very difficult, and as such we believe that replications would be necessary to further validate the results that we observed. It could be replicated on a similar sample, or on employees that just entered their first job. After all, we said that our results should apply just as well to this second group, but this does not mean a rigorous testing is not necessary. Any difference between our study and one that focuses on this second type of sample would teach us a lot.

The other limitation is the measure of performance. A measure with more heterogeneity would probably have been better. Besides, managers will tend to give a good grade to their interns to thank them for their efforts and in order not to get them into trouble with their school or university. Finding a better way of measuring performance for such complex tasks is no easy endeavor, but it could greatly increase the quality of such a study, and thus would make for very interesting research.

Nevertheless, in spite of those limitations, we still believe that this study was very informative, and we hope that decision-makers will consider this type of measure in their panel of evaluation tools.

Conclusion

In this study, we started by showing what affected satisfaction at work before exposing the shortcomings of the current method and showing what alternatives have been developed. Then, we introduced our hypotheses and the design that we put in place to test them. In the remainder of the paper, we presented our findings before discussing them.

We really believe that this method shows a lot of promise and could be a good complement to the current measures of satisfaction. We criticized them a lot, but they still have their own strength and purpose, so we are not saying that they should be abandoned. Rather, we believe that if they are combined with the method we proposed, and combined with the knowledge about what other elements of aspirations and evaluations influence satisfaction the most, they could be used to greatly enhanced not only the experience of interns, but also of new employees who just landed their first job. And if they are more satisfied, these workers will gladly return the favor by being more productive at their job.

Of course, satisfaction is only a small part of what influences productivity. Another important element is motivation. In the next chapter of this thesis, we will see how the Nudge approach can be used to improve motivation at work via a modern approach to the problem: gamification.

Chapter 4 – Nudge and motivation at work: the case of gamification

Introduction

Beside the current world debate on the crucial aspect of business competitiveness, the challenge of providing a fulfilling environment to employees has lately been a serious concern to major companies. Hence, motivation and engagement at work are increasingly seen as an effective way to bring productivity and sustainability to companies. This is why the study of emotions in the workplace, be they positive or negative, is nowadays drawing more attention than ever (Ashkanasy, Härtel, & Zerbe, 2000; Bono, Foldes, Vinson, & Muros, 2007).

In fact, happy employees are almost 20% more productive than unhappy ones (SGROI, 2015). But the sources of “unhappiness” in work could be diverse. In this study, we are focusing on repetitive and routine work tasks that, we think, are most of the time a cause of demotivation for employees. Often, monetary incentives are offered to counterbalance this lack of motivation, but it is legitimate to ask about the extent of their effectiveness. In that sense, the so called “crowding out effect” (Frey, 1997) shows us that financial incentives - being a major source of extrinsic motivation- have, at a certain level, the disadvantage to undermine intrinsic motivation and thus diminish a person’s efforts when performing a task, mostly due to impaired feelings of self-esteem or self-determination experienced by individuals.

Enhancing workers’ performance goes with the issue of fostering individuals’ engagement. In fact, 2017’ Gallup report showed that disengaged employees cost the U.S. \$483 billion to \$605 billion each year in lost productivity (O’Boyle & Harter, 2017, p. 19). Literature has also brought numerous evidences that a person’s favorable psychological state of mind correlates positively with its performance (Avey, Reichard, Luthans, & Mhatre, 2011; Richardson, Abraham, & Bond, 2012).

Hence, recent developments of continuously innovative management techniques provide a lot of solutions to engage employees in co-working, finding synergy and efficient cooperation in group tasks and providing, as much as possible, feelings of self-accomplishment and satisfaction when performing a task at work. But still, non-engaging tasks, like in a clerical job or repetitive service work, represent a hurdle to maintaining a suitable level of motivation and engagement.

Gamification is a new approach that we aim to test as an efficient alternative to monetary incentives for enhancing individuals’ engagement and performance. Gamification could be defined as “*the use of game design elements in non-game contexts*” (Deterding, Dixon, Khaled, & Nacke, 2011).

Conscious about a largely spread misconception on an existing dichotomy between “working” and “playing”, recent applications of gamification’s methodology try to show that those two activities are not incompatible. More than that, gamification proposes its methods to various fields including: education by implementing diverse technological contexts to foster enjoyable learning environments (Dicheva, Dichev, Agre, & Angelova, 2015; Kapp, 2012), well-being by influencing healthy behaviors using fitness apps (King, Greaves, Exeter, & Darzi, 2013; Lister, West, Cannon, Sax, & Brodegard, 2014), marketing by emphasizing experiential and playful contexts to consumers (Huotari & Hamari, 2012) and so on.

From this matter of fact, we have set, as the main purpose of this article, to evaluate gamification’s impact on motivation and job performance. Thereby, the main research question of this paper is the following:

“Could gamification, applied in work environments, have an impact on employees’ motivation and performance, in the case of simple and repetitive service tasks?”

To conceptualize the “service task”, we may take the example of employees working at a *Call Center* separately from other co-workers. Why? First, we need a type of task with a clear, quantifiable outcome that we could accurately link to the result of individual’s effort. If the task that we consider is too complex, then it would be difficult to determine whether people perform poorly because they did not try hard enough, or because they simply couldn’t handle the said task. Conversely, if the task is too simple, it would be difficult to generalize the results to workplace settings, where every day employees’ tasks tend to grow and become more and more arduous. We also consider solitary work because team work adds another layer of complexity that would make it very difficult to attribute our results to a single cause.

We try to grip a satisfying middle ground between simplicity and applicability by settling for “service tasks” that imply a clear binary outcome (you either succeed or you don’t, there is no in-between), and where the difficulty can vary randomly while, in the absolute, remaining simple and only requiring basic skills that could be learnt with no hassle. We also choose the designation “service task” because this terminology seems quite appropriate to describe this type of activity. For example, a customer support employee will receive calls from customers with issues. At the end, the outcome is binary (the issue is either resolved or not), clearly quantifiable (we can easily assess whether it was a success or not), customer issues can be basic or a bit more difficult to solve, and this type of activity is generally easy to learn. As a common consequence, generated by the nature of this kind of tasks, it can get repetitive and tedious quite quickly. Hence, the need for additional motivation relays that we offer to explore in this paper.

We propose to explore two types of gamification. In one case, we will offer virtual badges and medals to subjects to reward them for their work. This plays on extrinsic motivation. In the other case, we will build a game in which the subject will see a virtual town grow when he accomplishes more, which will provide a more intrinsic motivation by making the task seem more meaningful.

This paper will have the following structure: we will first explain the concept of gamification and its theoretical background, as well as discuss the general relationships between motivation and performance revealed in the joined literature of economics and psychology. Then, we will propose an economic model of gamification, to illustrate how it works in theory. Then, we will present our hypotheses and the experiment in which we propose to test them. Finally, we will present our results and discuss them.

Part 1 - Why Gamification?

Recent years have shown to what extent the game industry has been successful to address the engagement issue through a noticeable power to motivate people. According to (TechSci Research, 2019), the global gamification market was valued at \$ 6.8 billion in 2018 and is projected to grow at an impressive CAGR (Gross Annual Gaming Revenues) of 32% to reach \$ 40 billion by 2024 on account of growing demand for customer experience enrichment and improved engagement of employees. But actual researches tell us very little about how gamification could be put into practice in a working environment, even if there is a noticeable interest to address this challenge.

One of our first impressions when exploring the gamification concept was that literature doesn't agree about a standard definition of it. Continuous attempts to encompass the diverse uses of gamification resulted on some vague definitions of the term. One of the prominent definitions is proposed by (Seaborn & Fels, 2015, p. 17), where Gamification is presented as "*the intentional use of game elements for a gameful experience of non-game tasks and contexts. Game elements are patterns, objects, principles, models, and methods directly inspired by games*".

Symbolic rewards like badges and ranks given in colleges, businesses or military schools are not something new. Their uses have always been oriented to incentivize and also to reward people for performing a desired behavior or accomplishment. However, simple rewards used apart from game mechanics do not guarantee an inner process that boosts the rise of internal motivation, and therefore are excluded from being assimilated to gamification applications.

By departing from this idea of generating an inner virtuous process that could produce a willingness to act in a certain manner, emerges a particularly promising way to approach the study of gamification. This is from the use of motivation theories' perspective. In fact, extensive researches on the impact of rewards and different forms of incentives on motivation have already been investigated by the Self Determination Theory (SDT) (Ryan & Deci, 2000). But what could be said when considering the context of game applications? Also, little has been said about the economic origins of motivations. This is another point of interest that we are aiming to consider jointly with psychology's contributions.

Using gamification to Nudge performance: A behavioral economic interpretation

Performance is a challenging question to businesses that are facing an ever-increasing competition, forcing them to dig into new solutions to enhance the motivation of their employees. In order to reach this aim with gamification, we ask two questions: How could we implement gamification at work? And what effects could it have?

We think that the answer to these questions could offer valuable insights to the business sphere and also contribute to the enrichment of psychological and economic theories about the use of game elements on non-gaming contexts. But, first of all, we will clarify some of the terms that we will use.

What is gamification?

Since we are only interested in the concept of gaming, it is important to avoid any confusion between playing (*paidia*) and gaming (*ludus* in ancient Greek), as they represent distinct types of activities. (Caillois, 1961) was the first author to distinguish between the two concepts, claiming that playing involves **free-form, non-rule-based** and expressive actions in contradiction to gaming that involves a **rule-based** and **goal-oriented** form of playing.

More recently, in order to complete the previous explanation of the game concept, (McGonigal, 2011) exposed four critical features any game design must have to fit the precedent classification of *ludus* :

- Clearly defined **goals** that provide players with a purpose for playing the game;
- Consistently defined **rules** that represent the limitations and boundaries on how to achieve the given goals;
- A steady **feedback** system that guarantees that the goals can be reached if the game rules are respected;

- The **free will** of accepting participation in the game and thus following its rules to reach the goals.

Gamification versus Pointification

Numerous studies and applications of gamification's methods and effectiveness resulted in misunderstandings. Due to a lack of matured theoretical foundations and profuse applications among practitioners and researchers' communities, the study of gamification led to many contradictions and debates on its diverse mechanisms and applications. We wanted to clear out confusions that may arise between two distinct schemes that are gamification and pointification.

Pointification is considered, most of the time, as a "*subset of gamification*" (Massung, Coyle, Cater, Jay, & Preist, 2013). Firstly, named in 2010 by Margaret Robertson, a British game developer, in her blog "*Hide&Seek*", the concept of "Pointification" was brought as a criticism of an actual misconception and conflating between "points and games"²⁸.

Pointification solely relies on "Points, Badges and Leaderboards", also called the PBL triad (Pankiewicz, 2016) or "Point-based elements" (Kifetew *et al.*, 2017), while Gamification is considered as a more complex, dynamic and responsive mechanism. In order to grow more intrinsically valuable responses, gamification mindfully integrates immersive elements like challenges, characters, stories and virtual purposes (in our case: building a beautiful town).

Realizing how much individuals' involvement, intrinsic motivation and engagement could be important factors when pursuing a beneficial change in a person's behavior, this chapter also tries to understand if pointification and gamification have the same impact on individuals.

Human motivations at work: the economic view of incentives deepened by psychology?

Standard economics has been, for a long time, criticized on what some authors describe as the unrealistic assumptions of its foundational models. In fact, two major assumptions of a rational individual stand on the postulate that an agent: 1) has well-known and stable preferences overtime and 2) seeks permanently to maximize them. From the classical point of view of labor economics, in order to maximize employer's profitability, workers' wage should be based on their marginal productivity. Productivity being considered to be the key determinant of what a worker could claim.

²⁸ See Margaret Robertson, « Can't Play, Won't Play », Hide & Seek, October 6, 2010, accessed September 4, 2015, <https://hideandseek.net/2010/10/06/cant-play-wont-play/>

New Keynesianism allegations coming from well-known authors, like (Shapiro & Stiglitz, 1984; Solow, 1979; Yellen, 1984), came afterwards to contradict this postulate through what is commonly known as the “efficiency wage” hypothesis. This, in addition to reversing the previous relationship between wages and productivity, explains the causes of market-clearing deficiencies in the formation of wages in some markets causing unemployment. The major reason is the voluntary manipulation of wages by employers to keep them above the market equilibrium. Mainly, their motivations stand on the will to increase their employee’s productivity and reduce the costs induced by turnovers and shirking.

In parallel, the standard model of agency theory assumes that work is painful but increasing pecuniary incentives motivate increased levels of effort. Hence, no *extrinsic* incentives at all would result in the lowest possible level of effort.

Without contradicting the postulates of agency theory, we enquire about the possibility that employees who experience inner feelings of pride, self-esteem, or merely pleasure (forms of non-pecuniary incentive) will raise their efforts out of a utility-enhancing intrinsic motivation.

Linking the underlying psychological mechanisms of motivation to the neoclassical microeconomic theories of human behavior is still an under-explored issue. Self-Determination Theory (SDT) offers a new psychological approach to human social development and well-being and distinguishes between different levels of motivation (Amotivation, extrinsic and intrinsic motivation). This approach begins to be recognized as the most suited psychological methodology to complete the neo-classical inquiry of human motivations. In fact, from an efficiency wage theory perspective, motivation comes from receiving a wage above the competitive equilibrium without any changes in work conditions. By contrast, this study examines changes in motivation induced by changes in work conditions. Non-pecuniary changes of this sort substitute for pecuniary changes in wages by enhancing intrinsic motivation.

Motivation and performance at work

By definition, “the term **motivation** refers to factors that *activate, direct, and sustain* goal-directed behavior...**motives** are the “whys” of behavior - the needs or wants that drive behavior and explain why we do what we do. We don’t actually observe a motive; rather, we infer that one exists based on the behavior we observe” (Nevid, 2012, p. 288). To be more precise, (Nevid, 2012) distinguishes two types of motivation that he defines as following:

- The extrinsic motivation comes from outside of the individual and often involves external rewards such as trophies, money, social recognition or praise;

- Intrinsic motivations are motivations that originate from within (inside) the individual, such as trying to solve a puzzle purely for the self-gratification of it.

Preserving employees' motivation is a crucial way to sustain the future of any company. But this mission is more complicated than it sounds, especially when the work implies repetitive and tedious tasks. These categories of tasks are known to undermine intrinsic motivation. The frequently related causes are: unmet expectation causing boredom (Gkorezis & Kastritsi, 2017), resentment and a lack of self-esteem (McAllister & Bigley, 2002; Pierce & Gardner, 2004) generated when performing those tasks.

This study was designed to elicit new means of incentivizing repetitive work-tasks accomplished in seclusion and thus, in some way, neutralizing boredom and amotivation inherent to these constraints. We investigate whether this could be done by visual and playful manners to preserve motivation and performance at work without overusing financial compensations. The result could be qualified as a win-win situation where employers will not have to invest too much money to incentivize employee, while they would enjoy a more pleasant work experience.

Part 2 - An economic model of gamification in the workplace

Games seem to have a potentially transforming power that reduces constraints and some of the demanding or boring aspects associated with work. According to (Matallaoui, Hanner, & Zarnekow, 2017), gamification tends to change individual behavior by simply altering the context surrounding a decision, without resorting to monetary rewards or punishment.

We present below a choice-theoretic model of individual labor supply that makes a clear distinction between extrinsic motivation and intrinsic motivation. This simple formulation will enable us to compare the respective effects of pointification and gamification on a worker's effort and performance under two widely used incentive structures: fixed payment and piece rate. This model demonstrates in a conventional economic framework how extrinsic and intrinsic motivations interact with various incentives. In particular, it straightforwardly predicts the crowding-out effect of piece rate on pointification and gamification.

The psychological distinction between extrinsic motives and intrinsic motives closely parallels the economic distinction, initiated by (Robbins, 1935) between income and leisure. Income is an extrinsic motive and leisure is an intrinsic motive. You will not accept a job if you don't get paid for it. By contrast, leisure is defined in economics as any activity which an individual would do without being paid for it. The extrinsic-intrinsic distinction generalizes the income-leisure distinction. Given that any action must be motivated, an action's motive is extrinsic if the action would not be chosen without it.

If a salary provides an obvious extrinsic motivation for working, the quest for status is also an extrinsic motive, although a non-pecuniary one, if it is rewarded by a conspicuous signal like a medal. Alternatively, an action is intrinsically motivated if it does not require an extrinsic motive to be chosen. Hence, extrinsic motives must be observable but intrinsic motives need not be. Watching a movie is an observable type of leisure but listening to music while working may provide an unobservable component of leisure to work.

Let us now examine how pointification and gamification influence an individual's performance at work and interact with the pay scheme. Their impact is channeled through the two carriers of utility, wage (income and consumption) and leisure or, equivalently, work effort, which is the reverse side of leisure that relates directly to measurable performance at work. We begin with a few notations. We designate wage by w , non-pecuniary job amenities by Z , leisure time by l , and work effort by e , which are all non-negative variables. The worker's utility function may be written alternatively $U(w, Z, l)$ with $U'_w, U'_z, U'_l > 0$; or $U(w, z, e)$ with $U'_w, U'_z > 0, U'_e < 0$.

The first expression $U(w, Z, l)$ may be used to separate work time from breaks, which form the observable component of leisure at the workplace; and the second expression $U(w, Z, e)$ helps formulate the allocation of effort to performance. Indeed, assuming that labor is the sole factor of production, performance (output value) $\equiv y = qe$, which defines the individual's ability ($q > 0$). Pointification and gamification may be considered both as a specific job amenity Z . Pointification provides a status-like extrinsic motivation for allocating effort to work as the number of points, badges or trophies steadily increases conspicuously with work effort. Gamification contains an element of pointification too by associating the performance at work with the performance in a game, like a number of virtual houses built by the productive worker.

However, in contrast with the "cold" incentive provided by "points", building a town or winning a game is also a "hot" intrinsic motivation for work insofar it develops a feeling of competence and/or autonomy. Thus, we expect gamification to enhance utility more strongly than pointification insofar both amenities are non-pecuniary and equally conspicuous but a game seems more vivid than points.

Moreover, pointification and gamification are a special kind of job amenity: the number of points accumulates in proportion to performance at work. Hence,

$$Z = \zeta(z)qe \text{ with } \zeta(z) \geq 0, z=(\text{point, game}) \quad (1)$$

$$\text{with } \zeta(\text{game}) > \zeta(\text{point}) \quad (1')$$

Equation (1) clearly demonstrates that *pointification and gamification act like a virtual piece rate*. Consequently, their incentivizing effect will not be the same under a fixed payment scheme or a piece rate.

We study the interaction of factors of motivation with the following pay schemes:

- Fixed payment: $w = \bar{w}$; (2, fixed)
- Piece rate: $w = pqe$ with $0 < p < 1$ (2, piece),

where \bar{w} is the fixed wage and p is the constant piece rate.

We will put more structure on the worker's utility function, inspired by self-determination theory, by assuming that effort and rewards (in cash or in kind) are independent or complements in the production of utility whereas pecuniary and non-pecuniary rewards are substitutes. The first hypothesis means that the marginal disutility of effort diminishes at large with rising rewards: growing rewards provide an extrinsic motivation which activates the willingness to perform well, which is a standard assumption of efficiency wage theory. The second hypothesis means that the marginal utility of our specific job amenities diminishes with rising wages: receiving a virtual badge provides more pleasure and arousal to a low-paid worker than to a top executive, which is consistent with the psychological theory of crowding out of an intrinsic motivation (pleasure and arousal) with an extrinsic motivation (extrinsic rewards).

Fixed payment

In the fixed payment, the worker's utility function scheme takes the general form:

$$U(w, Z, e) \text{ with } U'_w, U'_z > 0, U'_e, U''_{ee} \leq 0, U''_{ew}, U''_{ez} \geq 0, U''_{wz} \leq 0 \quad (3)$$

To simplify the model's resolution and discussion, we choose the following utility function at work, given our emphasis on gamification:

$$U(w, Z, e) = w + \beta we + \zeta(z)qe(1 - \sigma w) - \frac{1}{2}ce^2(1 - \gamma\zeta(z)q), \text{ with } c, \beta, \sigma, \gamma > 0. \quad (3')$$

Compatibility of (3') with the conditions on the first and second derivatives of the utility function (3) require the following conditions: $1 - \sigma w > 0$, $1 - \gamma\zeta(z)q > 0$, $\beta - \sigma\zeta(z)q \geq 0$.

The optimal effort results from the maximization of utility function (3) or (3') under gamification constraint (1), budget constraint (2), and an effort constraint: $e \geq e_0 > 0$ (4), e_0 being

the minimum effort level that would prevail when no extra motivation is provided ²⁹. This generates four outcomes which can be compared with one another and with a control situation in which job amenities do not depend on performance. The latter situation is obtained when $\zeta(z) = 0$. Then, the corner optimum obtains under fixed payment: $e = e_0$, that is, the worker has no incentive to work hard and she makes the minimum effort. Since the purpose of the piece rate scheme is to provide an incentive to work harder, it can be assumed that the solution is interior with a piece rate ($e > e_0$), which requires: $pq > ce_0$. We make this assumption in the rest of our discussion.

Under the fixed payment scheme, the first-order condition (f.o.c.) of maximization of (3') is the equality:

$$\beta w + \zeta(z)q(1 - \sigma\bar{w}) - ce^*(1 - \gamma\zeta(z)q) = 0, \quad (5)$$

which determines the optimal effort:

$$e^* = \frac{\beta w + \zeta(z)q(1 - \sigma\bar{w})}{c(1 - \gamma\zeta(z)q)} \quad (6)$$

and performance qe^* . Equation (6) shows that, according to hypothesis (2'), workers would make more effort under gamification than under pointification. Effort is further enhanced by the ability (q), effort-wage and effort-points complementarity (β, γ), and impeded by the cost of effort (c) and the substitutability of pecuniary and non-pecuniary rewards (σ). Notice finally that ability interacts positively with gamification.

Piece rate

Under the piece rate scheme, the worker's utility function, takes the general form:

$$U(p, Z, e) \text{ with } U'_p, U'_z > 0, U'_e, U''_{ee} \leq 0, U''_{ep}, U''_{ez} \geq 0, U''_{pz} \leq 0, \quad (7)$$

and the simplified analog to (3') is:

$$\begin{aligned} U(w, Z, e) &= pqe + \beta pe + \zeta(z)qe(1 - \sigma p) - \frac{1}{2}ce^2(1 - \gamma\zeta(z)q) \\ &= (p(q + \beta) + \zeta(z)q(1 - \sigma p))e - \frac{1}{2}ce^2(1 - \gamma\zeta(z)q) \end{aligned} \quad (7')$$

Compatibility of (7') with the conditions on the first and second derivatives of the utility function (7) require the following conditions: $1 - \gamma\zeta(z)q > 0$, $q + \beta - \sigma\zeta(z)q \geq 0$.

(7') determines the optimal effort:

²⁹ e_0 may be interpreted as the collective norm of minimal effort required from workers for avoiding a rebuke or sanctions from supervisors, see (Akerlof & Yellen, 1990; Homans, 1954).

$$e^* = \frac{p(q+\beta)+\zeta(z)q(1-\sigma p)}{c(1-\gamma\zeta(z)q)} \quad (8)$$

Equation (8) shows that pointification and gamification act like a non-pecuniary piece rate if $1 - \sigma p > 0$. However, since workers, and especially low-paid workers, are certainly more attracted by monetary wages than by virtual badges or houses, we expect the impact of our non-pecuniary rewards to be much smaller than the money piece rate, i.e. $\zeta(z) \ll p$. This can be viewed as an instance of crowding out of an intrinsic motivation (our non-pecuniary rewards) by an extrinsic motivation (piece rate). Hence, we may conclude:

Proposition:

Under the fixed payment scheme, pointification and gamification are a virtual substitute for piece rate and exert thus an incentivizing effect on effort and performance. However, this effect is stronger for gamification than for pointification due to the stronger development of intrinsic motivation by gamification.

Under the piece rate scheme, the money piece rate crowds out the incentivizing effect of the virtual piece rates, pointification and gamification.

The individual's task-specific ability interacts positively with our non-pecuniary rewards and this effect is strongest with gamification and fixed payment.

Equations (6) and (8) demonstrate that the positive effect of gamification on effort and performance is robust to changes in the structure of the worker's utility function. By no means, gamification needs to be a complement to effort and/or a substitute to wage in the worker's utility function, as postulated by our interpretation of efficiency wage and self-determination theory. Indeed, the positive effect of gamification on effort and performance is obtained *in spite of* a potential crowding out of the intrinsic motivation provided by the playful aspect of games with the extrinsic motivation provided by a high wage. However, the positive effect of gamification on effort and performance is driven by the utility-enhancing, that is intrinsically motivating, effect of our non-pecuniary rewards (see equations (1) and (3)). To show this, assume that $\beta = \gamma = \sigma = 0$. Then, equations (6) and (8) boil down to:

$$e^* = \frac{\zeta_f(z)q}{c} \quad (6') \quad e^* = \frac{pq + \zeta_p(z)q}{c} \cong \frac{pq}{c} \quad (8'),$$

where ζ_p and ζ_f , with $\zeta_f < \zeta_p$, refer to 'fixed payment' and 'piece rate' schemes respectively.

Part 3 - Hypotheses

In this section, we will present the hypotheses that we will test in this work, but before we can proceed, we must first present one particularity of testing the effects of non-pecuniary rewards in a short experiment.

The two stages of non-pecuniary rewards: anticipation and awareness

In addition to the obvious monetary implications, there is one major difference between pecuniary and non-pecuniary rewards: one does not need to see the pecuniary rewards in action to understand their value.

When we announce to workers that they are going to be paid 10 € per unit produced, they can already decide what this reward is worth to them. They have a rough idea of what they can produce, so they can also get an approximation of how much they could earn during a day of work. Furthermore, they had ample time to experience the value of money before starting their job. As a consequence, workers could very well decide on the spot how much effort they want to do.

However, when we announce to workers that they are going to receive virtual badges or build a virtual town, this is completely meaningless to them as they have no experience of such things. At this point, the badges could be anything from highly-detailed pictures to very simple drawings, or even photos of real medals. Furthermore, they also do not have any idea of the increase in “quality” of the non-pecuniary rewards. Because of that, we cannot expect workers to see the value in gamification or pointification before exposing them to these rewards and letting them go through at least a few stages³⁰. It is only after having seen what the rewards look like, and what the progression curve (the increase in quality between each stage) looks like, that they can really start to appreciate the value of gamification or pointification. Only then can they decide if it is worth doing more effort to get more non-pecuniary rewards. In the context of a real-world application, where the test lasts for several weeks or months, this does not matter: the workers would only need a few hours to a few days to become well aware of how gamification works, and this is only a very short fraction of the experiment’s total time. However, when gamification is tested via a thirty-minute laboratory experiment, this initial “warm-up” time becomes very significant.

³⁰ In many applications of non-pecuniary rewards, including our own, workers are given these rewards as they reach a certain milestone, or “stage” as we call it here. For example, they would get a bronze medal for producing X units, then a silver medal for producing X more units after that, ... This is what we refer to in this paragraph.

In practice, this means that the motivation increase cannot be expected to happen as soon as the work session starts. There is first what we call an “anticipation” phase, where the subject is still wondering about gamification and is slowly finding out what it is about, and then an “awareness” phase where the subject is fully aware of how gamification works and what to expect. And it is only in this second phase that gamification or any other form of non-pecuniary rewards would start having an effect³¹. As a consequence, we will evaluate the effect of gamification on productivity only in the “awareness” phase. Our method for determining the point when one phase ends and the other begins is empirical and will be detailed in the results section.

The hypotheses

The model presented in section 3 allows us to draw a number of hypotheses.

Hypotheses H1 to H5 are formulated in the context of a fixed salary. The theoretical literature and our model show that monetary incentives can crowd-out intrinsic motivations. The direct consequence of this phenomenon is that it could be difficult to observe the effects of non-pecuniary rewards when a piece rate is already implemented. In order to be able to correctly evaluate their effects, we need to observe them without any interference.

H1: Gamification and pointification improve worker productivity in the awareness phase.

According to our model, gamification and pointification will both increase a worker’s effort because they constitute a form of non-pecuniary reward. And because of what we explained in section 4.1, we only expect this effect to occur in the awareness phase. This leads us to our second hypothesis.

H2: Gamification and pointification do not improve worker productivity in the anticipation phase.

We do not expect the mere mention of a gamification system to have any effect on worker productivity.

H3: Gamification increases intrinsic motivation compared to a control group and to pointification.

Since gamification is more vivid and playful than pointification, we expect it to have better effects on intrinsic motivation than the control and pointification since the latter only plays on non-pecuniary, extrinsic motivation.

H4: Gamification and pointification decreases the worker’s willingness to take a break.

³¹ Of course, one could simply show the pictures and rewards to the worker in advance, but doing so would spoil the surprise, which is a big part of the appeal.

Since both treatments should increase a workers' motivation, we expect that they would choose to diminish their leisure time.

H5: The individual's task-specific ability interacts positively with gamification.

For the sake of this experiment, we will make the assumption that an individual has an initial level of skill for a given task, before any learning effect or motivator comes into action. Our model showed that this ability interacts positively with gamification. Indeed, everything else being equal, one additional unit of effort from a highly skilled individual will produce better results than one additional unit of effort from a less skilled individual.

H6: Under a piece rate, gamification and pointification do not increase productivity and intrinsic motivation, and they do not decrease the worker's willingness to take a break.

If there is indeed a crowding-out effect, then implementing gamification or pointification when a piece rate is already in place should have negligible effect on any of the variables mentioned above.

Part 4 - Experimental design

We have chosen to study gamification's impact through a lab experiment because it offers, from a behavioral economics point of view, a controlled and isolated environment, that, we think, is critically important when pursuing the purpose of avoiding situational factors. The latter may have a significant impact on the effect of various types of incentives and gaming frameworks.

Procedure and design of the study

The task

Our goal was to represent a service task (as we conceptualized it earlier). This task should be simple and repetitive, of randomly variable difficulty, and require some problem-solving skills. To achieve these features, we used a word-building task: we first took the 226 most common five-letter words in the French language (the lab experiment took place in Paris and all subjects were fluent in French), then randomly shuffled the letters before presenting the result to the subjects. We refer to these as "word sequences". Then, we asked the subject to try and build a word with the letters that they saw on the screen. Hence, the task was repetitive and relatively simple because all they did was building common words. However, the difficulty still had a bit of variability. In fact, even in our list, some words were more common than others, and certain random arrangements of letters made it

easier to guess the answer than others. The fact that we limited ourselves to 5-letter words does bound the difficulty so the randomness is not an issue.

The word sequences appeared sequentially on the screen. Each subject had up to 60 seconds to build a word out of a given sequence. The fact that we built each sequence out of a real word ensured that there was always at least one possible answer, but sometimes there were multiple correct answers. In that case, the subject could choose any of these answers as the correct answer before going to the next word. It didn't matter which word was provided, as long as it was correct.

If subjects failed to find a word in 30 seconds, they could skip to the next one. They were not allowed to skip before 30 seconds because we did not want them to mindlessly skip every word sequence until they stumbled upon an easy one. If they did not find anything after 60 seconds, the program would automatically skip to the next word.

The process

First, the subjects were told that they would take part in an experiment where they were expected to perform a simple real-effort task.

It started with a 5-minute training session. The task was explained to the subjects and they were asked to build as many words as possible during these 5 minutes. There was no incentive or gamification at this stage, they were simply told to take this exercise seriously as it would be useful later on.

After the training session, they were told that they would have to perform this task for 30 minutes. This is what we call the "work session" of the experiment. After every ten words (no matter how many of them the subject actually managed to find), a "transition screen" showed how many sequences were successfully turned into correct words. This screen was implemented in each treatment, in a way that will be detailed in the next section. During the work sessions, they were allowed to take up to 5 minutes of break. To do so, they just had to press a button that was constantly available, except during the transition screen. It would take them to an alternative screen that would display a relaxing video.

After the work session, subjects were asked to answer a series of questions. First, they had to fill the French version of the Ten Items Personality Inventory (Gosling, Rentfrow, & Swann Jr, 2003), a short version of the Big Five personality inventory which measures five dimensions (extraversion, agreeableness, conscientiousness, neuroticism, and openness) via ten items on a scale of 1 (not at all) to 7 (totally). Then, they were asked to fill the French version of the Intrinsic Motivation Inventory

(E. Deci & Ryan, 1985). It evaluates the intrinsic motivation of the subjects for the task that we asked them to do via four dimensions (feeling of competence, pressure, feeling of autonomy and enjoyment), each evaluated via five items on a scale of 1 (not at all) to 7 (totally). Finally, they were asked to provide a few more information about themselves: gender, age, diploma, the age at which they started learning French and the time they spent playing video games each week. Once they were done, they could collect their earnings. In addition to what they earned for the task (which depended on the treatment), they all earned 5 € for their participation.

The treatments

We did a 2*3 between-subjects design: two payment treatments and three non-pecuniary-reward treatments, for a total of six different treatments. Table 1 describes each treatment and how we refer to them from now on.

Table 32 – The treatments of the experiment

| Non-pecuniary reward Payment scheme | Control Group: no reward | Pointification: “Badges” | Gamification: “Town Building” |
|--|---------------------------------|---------------------------------|--------------------------------------|
| Fixed payment: 8 € for the work session | Fixed control | Fixed badges | Fixed town |
| Piece rate: 0.10 € per correct word in the work session | Piece control | Piece badges | Piece town |

In the fixed control group, the transition screen only informs the subjects of how many words they completed thus far. In the piece control group, it additionally informs the subjects of how much money they earned so far. In the other treatments, gamification is implemented exclusively via this transition screen.

The “badges” treatment is our implementation of pointification. In this treatment, subjects were told that they would earn virtual badges and rewards based on their performance. We explained to them that these would not affect their earnings, that they were purely symbolic and only lasted for the duration of the experiment (they knew that they could not take any real badge home). On the transition screen, an image appeared depending on the current number of correct words found by the subjects. At first, when they had found less than 10 words, it displayed an empty trophy case. Every 10 words found, a new medal or trophy appeared in that trophy case. There was a total of 20 pictures, which corresponded to how many “levels” the subject could unlock. In each

picture, a new badge or trophy appeared, and they were designed so that the rewards kept looking better and better. Once the subject had found 200 correct words, this was the maximum and she could not see new pictures anymore. Only about 1% of the subjects reached that score. The figure below shows the pictures we used for the experiment:

Figure 14 – The pictures of the “badges” treatment (first reward on the left, complete reward set on the right)



The “town building” treatment is our implementation of gamification as we defined it earlier. In this treatment, subjects were told that their effort would lead to the construction of a virtual town: with every 10 words built, their town would grow bigger and look nicer. They were explicitly told that this was just a game, that the advancement of the town has absolutely no real world meaning or consequence, and that there was no reward for building a bigger and nicer town. Apart from the visual and theme difference, it functioned in exactly the same way as the “badges” treatment: they saw a new picture every 10 correct words, there was a total of 20 pictures, and they were designed in such a way that they looked better and better.

Figure 15 - The pictures of the “town building” treatment (first reward on the left, complete reward set on the right)



The “town building” treatment and the “badges” treatment differ in major ways. Individuals obtain what is essentially an extrinsic reward in the “badges” treatment, since badges have always

been used as a token of respect and recognition. On the other hand, individuals may feel competent and proud of building something in the “town building” treatment.

Moreover, finding 10 new words meant getting between 2 and 3 new buildings in the “town building” treatment whereas the same performance was rewarded with a single new badge in the “badges” treatment. People could feel like they accomplished more in the “town building” treatment than in the “badges” treatment.

Therefore, the difference between the two treatments, while being purely visual and not functional, could still have an impact. Indeed, we expected the “badges” treatment to affect the extrinsic motivation of individuals: they would work harder to earn badges, just like they would want to earn more money. As for the “town building” treatment, we expected it to impact the intrinsic motivation as well: it would make the task feel more meaningful, livelier, and more enjoyable.

The gamification and pointification both remained very basic and similar to one another in a functional sense. Indeed, we wanted to investigate whether even the most straightforward and easy-to-implement forms of gamification could have an effect, and we also wanted to see whether a simple change in visual theme could be the difference between gamification and pointification.

The sessions and participants

We conducted a laboratory experiment during May 2018 via a total of 12 sessions (2 sessions for each of our 6 treatments) with a goal of 20 participants per session. Prior to these sessions, we conducted a single calibration session. This allowed us to evaluate the average performance of subjects during the work task when subjected to monetary incentives of 0.10 €. We then chose a fixed payment of 8 € so that the average earnings were approximately the same no matter how they were paid.

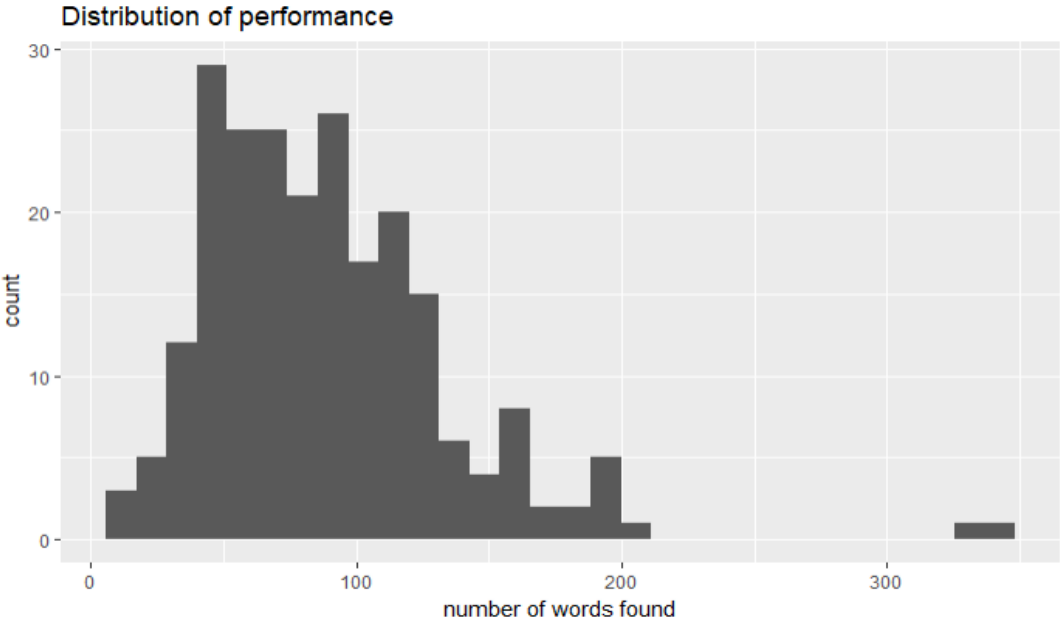
All participants were recruited randomly from the database of the website of L.E.E.P. (Laboratoire d’Economie Expérimentale de Paris). Subjects were randomly assigned to one of the 12 sessions. All participants received an e-mail invitation containing the location, date and time of the experimental session. They did not know the topic of the study prior to entering the laboratory.

Part 5 - Results

We gathered data from 228 participants, of which 52% were females and 77% were students (mostly in social sciences), with a mean age of 24 years. On average, they spent 2.6 hours per week on video games, with 58% reporting not playing at all. 80% of participants were native French speakers. Neither time spent on video games nor native language turned out to be significant in the experiment.

Before testing the implications of the model on the data, let us take a first look at the performance of our subjects in the experimental task. The figure below shows that the distribution of performance in the work session is skewed to the right, which is partly due the presence of two outliers who found all the words and whose performance was more than 5 standard deviations higher than the mean.

Figure 16 - Distribution of performance in the work session (outliers included)

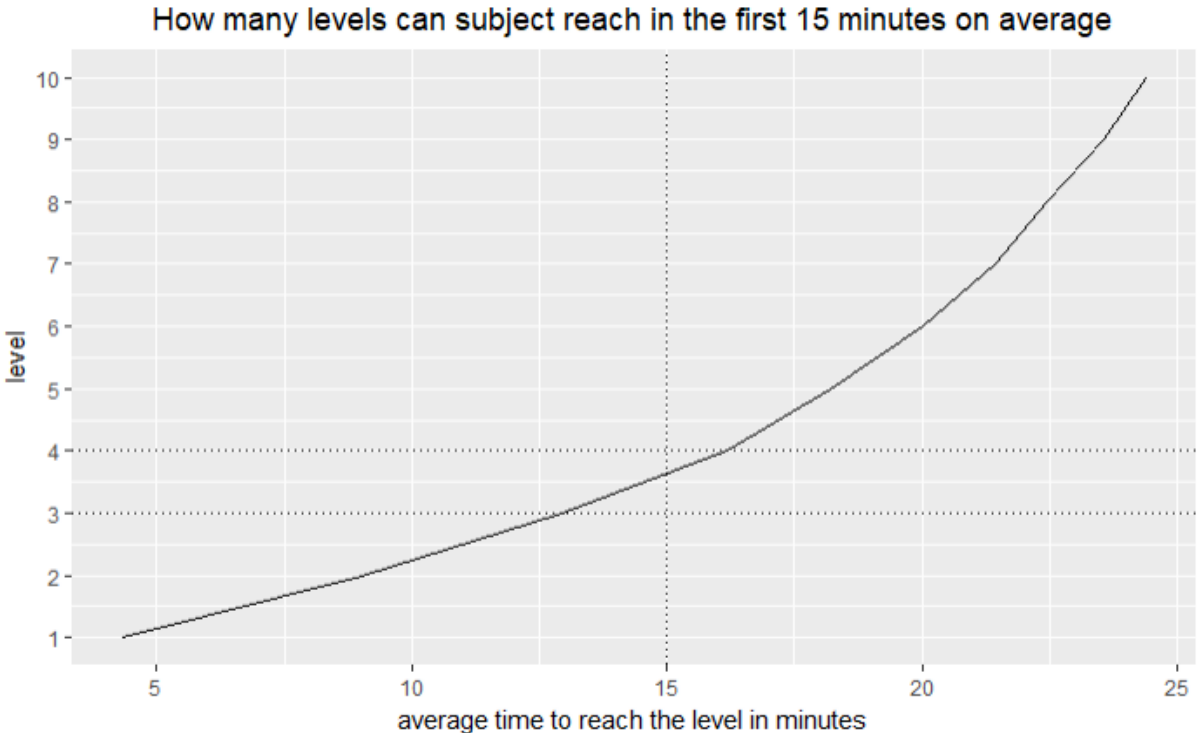


The figure below depicts the average time needed to find ten more words. On average, it took subjects a little less than 5mns -4'21" to be accurate- to find the first ten words (level 1) and visualize their first badge or house. During this initial period, subjects were not aware of what those badges or houses looked like. The figure suggests that a majority had seen at least 2 transition screens after 10 minutes and three transition screens after 15 minutes, which is just enough to start grasping the extent of the gamification and know what to expect later on.

In order to simplify the analysis of productivity over time, we cut the work session into six periods of five minutes each, and measured the number of words found in each period. Hence, given what we showed in the previous paragraph, we chose to consider that the awareness phase starts at fifteen minutes. This point is very convenient for the analysis because both phases will be of the

same size, it will give us three periods' worth of data for the awareness phase, and the fact that both the median and average subject will have seen three levels at this point means they will already be well aware of how gamification works.

Figure 17 - Average time required to reach a given level, in minutes



We tested the hypotheses using Student tests (when not specified otherwise, p-values reported in this section come from Student tests), OLS models and Wilcoxon rank sum tests (when the data was not normally distributed). The two outliers that appeared in figure 2 have been removed from the results presented below to increase the significance of the treatment effects³², reliably use the Student tests and compare means, which are common and easy to interpret.

Effect of gamification on performance

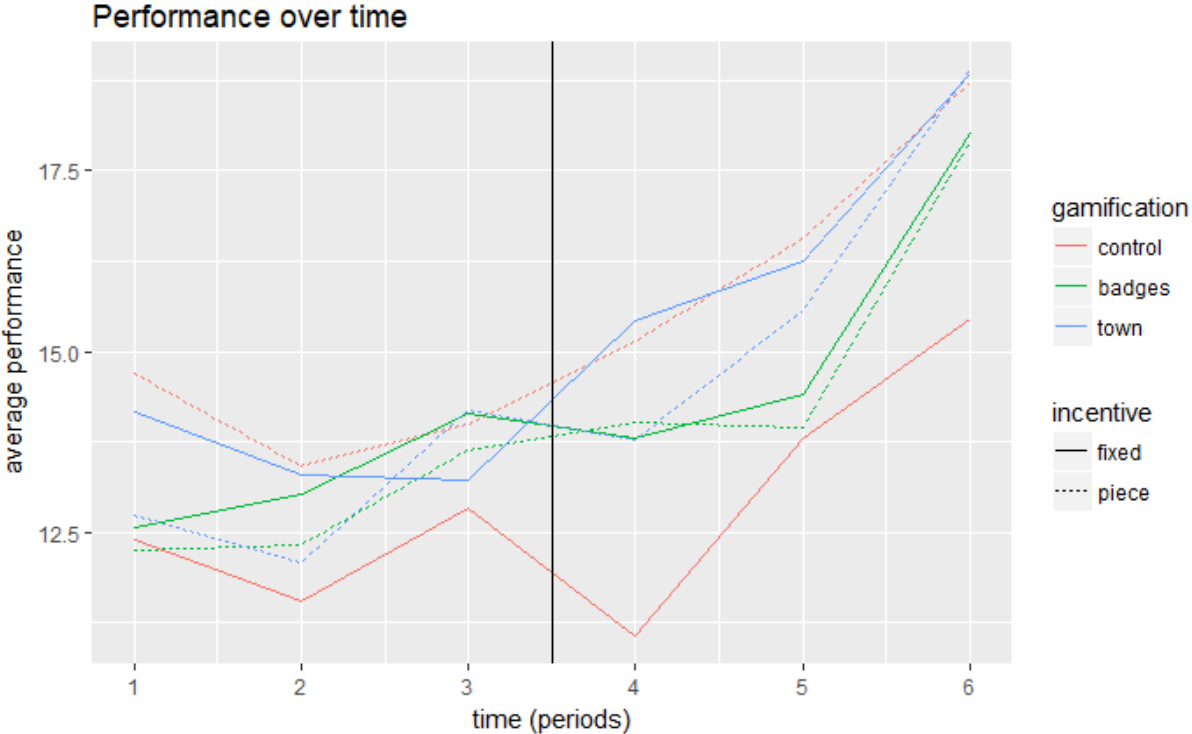
The figure below shows how performance evolves over time in each of the treatments. As explained earlier, we divided the experiment into six 5-minute periods, as shown on the x axis. The average performance at period 1 corresponds to the average performance during the first 5 minutes,

³² If we keep the two outliers, then the t-tests reveal no treatment effect on performance at the 5% level for gamification and piece rate in the anticipation phase and in the awareness phase. However, Kruskal-Wallis tests reveal that there is a significant difference at the 10% confidence level in the distribution of performance between “fixed control” and “fixed town” (p-value = 0.09701), and between “fixed control” and “piece control” (p-value = 0.07236). This suggests that there is, indeed, a difference in the treatments. Besides, removing the outliers had no effect on the treatment effects for intrinsic motivation and breaks.

the average performance at period 2 corresponds to the average performance between 5 minutes and 10 minutes, and so on. The vertical line separates the two phases: the three periods on the left of that line show the performance during the anticipation phase, while the three periods on the right of that line show the performance during the awareness phase.

Two things stand out on this figure. First, it seems that the speed at which performance increases during the awareness phase differs between treatments. To verify this, we conducted a series of Student tests of differences, using the difference between performance in period 6 and performance in period 4 as a simple measure of speed. We found no significant differences between treatments (for example, we obtained p-value = 0.398 between “fixed town”, which had the lowest speed, and “piece town”, which had the highest speed). Second, the potential treatment effects seem stronger in the awareness phase. We will investigate this more carefully in the remainder of this subsection.

Figure 18 - Performance over time per treatment



Effect of gamification on performance during the awareness phase

The table below shows the distribution of performance across treatments during the awareness phase.

Table 33 - Performance distribution across treatments in the awareness phase

| Treatment | Average | Standard deviation | P-value |
|---------------|---------|--------------------|---------|
| fixed control | 40 | 22 | - |
| fixed badges | 46 | 23 | 0.2637 |
| fixed town | 51* | 25 | 0.0613 |
| piece control | 50* | 23 | 0.0648 |
| piece badges | 46 | 23 | 0.2905 |
| piece town | 48 | 22 | 0.1261 |

*: significantly higher than "Fixed control" at the 10% significance level according to a Student test

We can see that only two treatments managed to improve productivity compared to a treatment with no pecuniary or non-pecuniary rewards: the gamification treatment, and the piece rate treatment.

The effect of the piece rate is unsurprising and very common. However, it is very interesting that the effect of gamification is of the same size as that of piece rate, which seems to validate the idea that non-pecuniary rewards can be equivalent to pecuniary rewards in terms of productivity increase... but only when implemented in the correct way. Indeed, pointification fails to increase productivity. This suggests that non-pecuniary rewards should focus on intrinsic motivation rather than extrinsic motivation, and that monetary rewards remain the best way to improve extrinsic motivation.

Therefore, we partially accept H1: gamification improves worker productivity in the awareness phase, but pointification does not.

We can also see that none of the treatments that combine pecuniary and non-pecuniary rewards manage to significantly improve performance. We will come back to this when we test for the presence of a crowding-out effect.

Finally, we built an OLS model to predict performance during the awareness phase using treatments and personality, with gender and age as control variables. The results are displayed in the table below. This model confirms the results of the Student tests. It shows that adding the "town" gamification to a situation where there was neither gamification nor monetary incentives increased the average number of words found by approximately 9.

Table 34 - Regression of performance on treatment and individual variables

| Performance in awareness phase | | |
|--|------------------|----------------|
| <i>Predictors</i> | <i>Estimates</i> | <i>p-value</i> |
| (Intercept) | 54.73 | 0.001*** |
| fixed badges | 4.63 | 0.403 |
| fixed town | 9.38 | 0.09* |
| piece control | 10.04 | 0.075* |
| piece badges | 5.71 | 0.31 |
| piece town | 8.11 | 0.144 |
| conscientiousness | -2 | 0.239 |
| extraversion | 0.08 | 0.956 |
| neuroticism | -1.17 | 0.407 |
| openness | 1.75 | 0.332 |
| agreeableness | -0.75 | 0.662 |
| age | -0.23 | 0.562 |
| man | -5.05 | 0.126 |
| Observations | 226 | |
| R ² / adjusted R ² | 0.050 / -0.004 | |

Understanding the effect of “fixed town” on performance

We just showed that gamification, when implemented in such a way as to harness intrinsic motivation rather than extrinsic motivation, can have a significant effect on performance. We also showed that the “fixed town” gamification had a powerful effect on the feeling of competence. Does that mean that it is this increase in feeling of competence, an element of intrinsic motivation, that led to that increase in performance? The table below shows the results of a linear regression on treatment and intrinsic motivation.

Table 35 - Linear regression on treatment and intrinsic motivation

| Performance in awareness phase | | |
|--|------------------|----------------|
| <i>Predictors</i> | <i>Estimates</i> | <i>p-value</i> |
| (Intercept) | 25.98 | 0.013** |
| fixed badges | -1.31 | 0.822 |
| fixed town | 1.42 | 0.807 |
| piece control | 6.35 | 0.28 |
| piece badges | 6.35 | 0.27 |
| piece town | 5.37 | 0.353 |
| interest | 0.72 | 0.61 |
| pressure | -0.39 | 0.754 |
| choice | -2.86 | 0.074* |
| competence | 7.44 | <0.001*** |
| Observations | 226 | |
| R ² / adjusted R ² | 0.144 / 0.108 | |

We can see that, when adding intrinsic motivation, the significant effect of the “fixed town” treatment disappears. On the other hand, “competence” has a very strong effect on performance. This seem to suggest that it was, indeed, the increase in intrinsic motivation that led to the increase in performance in the “fixed town” treatment. Seeing the visual rewards made them feel better about themselves, which made them want to do more.

One might also think that the increase in “competence” is just the result of an increase in performance: if people perform better, they should naturally feel more “competent”, and not the other way around. If that were the case, then we should expect no difference in “competence” between “fixed town” and “piece control”. We verified this with Student tests. There is no difference in performance between those two treatments in the anticipation phase (p-value = 0.757) and in the awareness phase (p-value = 0.9844). However, “competence” is significantly higher in “fixed town” than in “piece control” (p-value = 0.04717). This shows that an increase in performance does not necessarily generate an increase in the feeling of competence.

To conclude on this subsection, gamification can be effective, but it has to be implemented in such a way that it increases people’s feeling of competence. As such, subjects need a little time to fully appreciate it, and it must not be implemented along a piece rate.

Effect of gamification on performance during the anticipation phase

The table below shows the distribution of performance across treatments during the anticipation phase.

Table 36 - Performance distribution across treatments in the anticipation phase

| Treatment | Average | Standard deviation | P-value |
|---------------|---------|--------------------|---------|
| fixed control | 37 | 22 | - |
| fixed badges | 40 | 20 | 0.5503 |
| fixed town | 41 | 19 | 0.4131 |
| piece control | 42 | 20 | 0.2858 |
| piece badges | 38 | 20 | 0.7663 |
| piece town | 39 | 15 | 0.6143 |

As expected, gamification does not have any effect in the anticipation phase. We also tested whether the simple replacement of the fixed salary by a piece rate could have an effect on performance, but the difference between “fixed control” and “piece control” is not significant (p -value = 0.2858). This shows that, in the anticipation phase, absolutely no treatment has any impact on performance. The reason behind this phenomenon is simply that it takes time to become comfortable with the task, even though we first had a trial session. In that case, a lack of performance may simply indicate that the subject is still learning, and not that she lacks motivation. Past this initial stage, the task becomes easier for the subject, and that is when it is important (and impactful) to introduce motivators so that it does not become boring. One takeaway from this result is that incentives and gamification yield more results on people who are familiar with their job, as opposed to people who are still learning it.

These results allow us to accept H2: gamification and pointification do not improve worker productivity in the anticipation phase.

Effect of gamification on intrinsic motivation

To test this hypothesis, we used the Intrinsic Motivation Inventory, or IMI (Deci & Ryan, 1985). We performed four linear regressions, with the four dimensions of intrinsic motivation as dependent variables, as this method allowed us not to multiply Student tests and to avoid a multiple comparison error. We chose not to perform an ANOVA because we want to know where the differences are, not just if there are any.

In each of the regressions, the baseline treatment is “fixed control”. The results are indicated in the table below.

Table 37 - Regressions of intrinsic motivation components on treatment

| Variables | Pressure | | Interest/Enjoyment | | Feeling of Autonomy | | Feeling of Competence | |
|-------------------------|---------------------|------------|-----------------------|------------|-----------------------|------------|-----------------------|------------|
| | B | P-value | B | P-value | B | P-value | B | P-value |
| (intercept) | 3.2914 (0.2383) | <2e-16 *** | 4.90952 (0.25135) | <2e-16 *** | 4.67857 (0.20641) | <2e-16 *** | 3.89143 (0.22102) | <2e-16 *** |
| fixed-badges | -0.1076 (0.3324) | 0.7464 | -0.08070 (0.35063) | 0.818 | -0.07722 (0.28794) | 0.818 | 0.47073 (0.30831) | 0.1282 |
| fixed-town | 0.2586 (0.3263) | 0.4289 | -0.13452 (0.34417) | 0.696 | -0.11607 (0.28264) | 0.696 | 0.68857 (0.30264) | 0.0239 ** |
| piece-control | 0.5475 (0.3346) | 0.1033 | -0.10860 (0.35298) | 0.759 | 0.13393 (0.28988) | 0.759 | 0.11968 (0.31038) | 0.7002 |
| piece-badges | 0.7701 (0.3282) | 0.0199 ** | -0.11038 (0.34623) | 0.750 | -0.05678 (0.28433) | 0.750 | -0.17861 (0.30444) | 0.5580 |
| piece-town | 0.7034 (0.3282) | 0.0332 ** | 0.00928 (0.34623) | 0.979 | 0.10989 (0.28433) | 0.979 | -0.05553 (0.30444) | 0.8554 |
| R ² | 0.055 | | 0.001445 | | 0.00606 | | 0.05506 | |
| Adjusted R ² | 0.03352 | | -0.02125 | | -0.01653 | | 0.03358 | |
| F | 0.03352 | | 0.06367 | | 0.2683 | | 2.564 | |

Signif. codes: 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.'

We see that no treatment increased the interest/enjoyment for the task. To find an explanation for this lack of effect, we computed the median interest level across all treatments (we averaged their answers to the seven questions, all on a scale from 1 to 7) and obtained a median of 5.17. Since this number is already high, it could be that the task was not found especially tedious, and so there was little margin of improvement on that aspect. We conclude that gamification has no effect on interest for the task. Similarly, the treatments did not affect the feeling of autonomy. Gamification also failed to decrease pressure. Even worse, pressure increased when gamification and piece rates were implemented at the same time. We can also notice that the “piece control” treatment almost increased pressure at the 10% significance level (p-value = 0.1033). This means that piece rate probably increases the pressure on people, and having both piece rate and gamification definitely does increase pressure compared to a treatment with no rewards at all. Having all of these incentives together can be overwhelming to subjects. However, the “town-building” gamification significantly increased the feeling of competence, keeping other incentives constant (p-value = 0.0239). The fact that there is no significant difference in feeling of competence between “fixed control” and “piece control” even though performance is different between these treatments proves that this is a clear gamification effect, and not merely a consequence of the fact that subjects performed significantly better in “fixed town”.

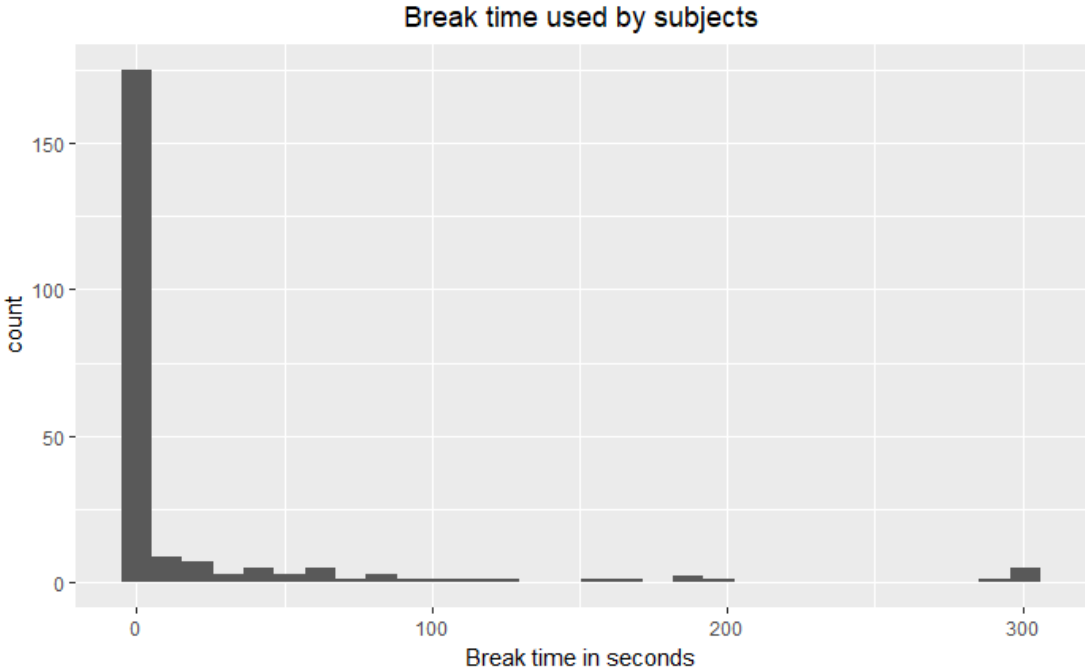
We also tested the differences in feeling of competence between “fixed town” and “fixed badges”, and we found it to be insignificant (p-value = 0.8675). This means that, while pointification fails to make subjects feel significantly more competent (although the p-value is close to the 10% significance level), it is not significantly worse than gamification.

In the end, we can partially accept H3. One dimension of intrinsic motivation, the feeling of competence, is improved by gamification compared to the control group. However, gamification is not significantly better than pointification in that regard.

Effect of gamification on break time

Overall, the “break” function of the experiment was ignored by most of the subjects, with 77% of them never using it. The figure below shows the distribution of the break time that was taken by subjects across all treatments.

Figure 19 - Distribution of break time



The distribution being clearly non-normal, we used Wilcoxon rank sum tests to verify H4. The results are displayed in the table below:

Table 38 – Break time per treatment

| | Average | SD | p-value |
|---------------|---------|-------|---------|
| fixed control | 22.54 | 65.16 | - |
| fixed badges | 18.59 | 53.03 | 0.7795 |
| fixed town | 37.1 | 77.72 | 0.3806 |
| piece control | 2.94* | 12.62 | 0.0889 |
| piece badges | 14.03 | 50.34 | 0.5348 |
| piece town | 21.26 | 56.66 | 0.9284 |

*: significantly higher than “Fixed control” at the 10% significance level according to a Wilcoxon rank sum test.

The only significant result is that people took less break time under a piece rate with no non-pecuniary rewards than when there were no rewards at all, which goes to show that monetary incentives still remain a good, albeit imperfect, motivator.

In light of these results, we can reject H4. Neither gamification nor pointification encouraged people to take less break time. One should remember that the task was easy and the work session only lasted for thirty minutes, which means the subjects probably just did not need a break. Furthermore, all they could do during the break was to watch a short relaxing clip, so many people probably thought that it was not worth their time.

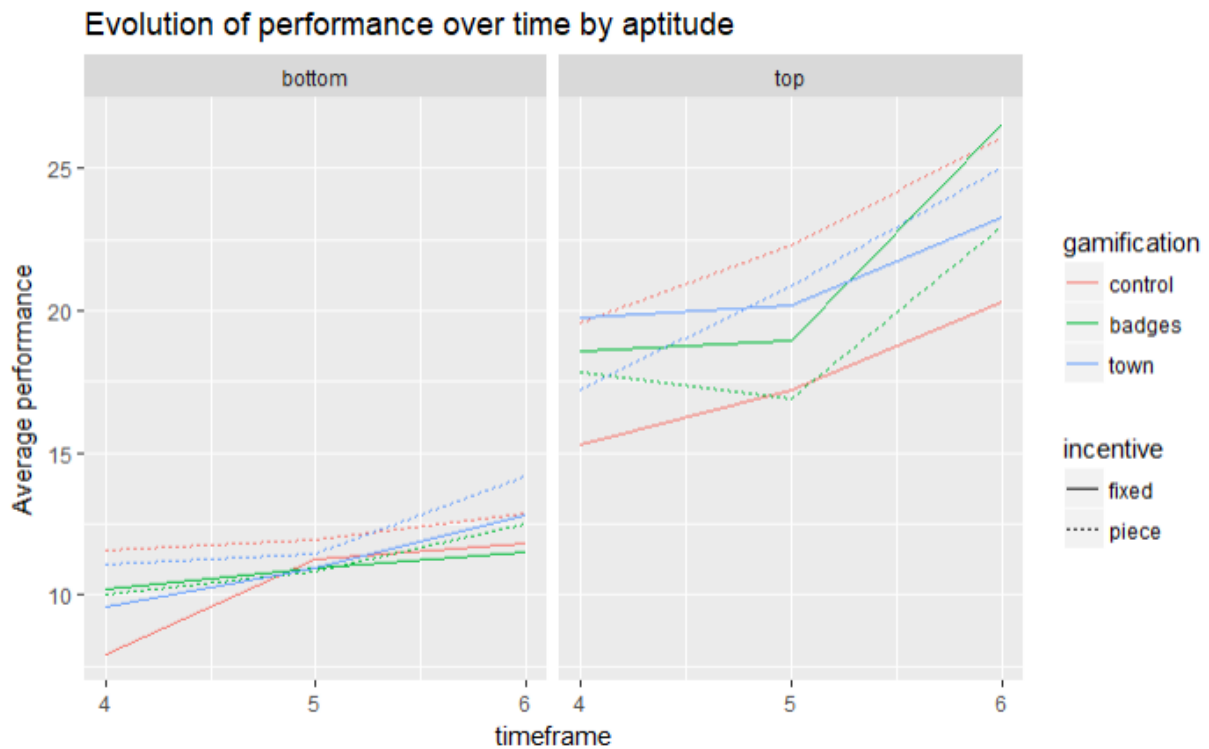
The interaction between gamification and ability

In our model, we showed that gamification interacted positively with skill. Individuals that have a better innate ability for the task should see a better productivity increase when exposed to non-pecuniary rewards.

In order to test this hypothesis, we divided the subjects into two groups according to their initial ability: the “bottom” group is made of the 50% of subjects who had the lowest score in the trial task, while the “top” group is made of the 50% of subjects who had the highest score in the trial task. The trial task consisted in working on the task for 5 minutes, without any break or transition screen, and it was implemented in exactly the same way for all treatments: without any form of pecuniary or non-pecuniary rewards. This makes it a good measure of the innate ability of subjects for this particular task.

Figure 9 shows a side-by-side comparison of how performance evolved during the awareness phase of the experiment in each treatment, for each of the two groups.

Figure 20 - Comparing the performance over time of the bottom performers and the top performers



A first look at the plot suggests that there is no difference between the treatments for the bottom group, while it would appear that non-pecuniary rewards improve performance for the top performers when they are paid a fixed salary.

We ran several Student Tests of differences between the “fixed control” group and the treatments to verify this intuition. The results are displayed in the table below:

Table 39 – Difference in performance between treatments for each group

| Treatment | Average performance of the “bottom” group in the awareness phase (p-value) | Average performance of the “top” group in the awareness phase (p-value) |
|---------------|--|---|
| Fixed control | 30.95 | 52.8 |
| Fixed badges | 32.67 (p = 0.7067) | 64.06 (p = 0.1629) |
| Fixed town | 33.35 (p = 0.6543) | 63.28 (p = 0.1693) |
| Piece control | 36.4 (p = 0.2871) | 67.94* (p = 0.0567) |
| Piece badges | 33.37 (p = 0.6321) | 57.75 (p = 0.5291) |
| Piece town | 36.73 (p = 0.2149) | 63.12 (p = 0.2085) |

*: significantly higher than “Fixed control” at the 10% significance level according to a Student test

The tests confirmed that there was no difference for the bottom performers, but they could not confirm that there was a difference for the top performers. We did observe that only the top performers were sensitive to pecuniary rewards. However, given what we can see on the graph, one can assume that the lack of significance is mainly due to a small sample size. For example, we only have 15 subjects in the “fixed control” group and 23 in the “fixed town group”.

Nevertheless, we have to reject H5. In this experiment, gamification does not interact with task-specific ability.

The effects of gamification under a piece rate

So far, we have mainly conducted tests of differences between fixed control and the different treatments. In the table below, we put the results of Student tests of differences between “piece control” and the two treatments that combined pecuniary and non-pecuniary rewards on key variables.

Table 40 - Tests of differences between “piece control” and the other treatments with piece rate

| | Performance (awareness phase) | Break time | Competence | Choice | Interest | Pressure |
|---------------|----------------------------------|------------|------------|--------|----------|----------|
| piece control | 42 | 2.94 | 4.01 | 4.81 | 4.8 | 3.84 |
| piece badges | 38 | 14.03 | 3.71 | 4.62 | 4.79 | 4.06 |
| piece town | 39 | 21.26* | 3.83 | 4.78 | 4.92 | 3.99 |

*: significantly higher than “piece control” at the 10% significance level according to a Student test.

The only significant difference is in break time, where people in the gamification treatment take more break time than those in the treatment with only the piece rate (p-value = 0.0559). After further investigation, we noticed that this difference was entirely driven by a single subject who took 300 seconds of break in the “piece town” treatment³³. If we remove him from the analysis, there are no differences in the average break time between the two treatments.

Interestingly, while we observed in section 5.3 that treatments with both pecuniary and non-pecuniary rewards yield a significant increase in pressure felt, we do not observe this phenomenon in this table. This suggests that the main driver of pressure was the piece rate, and not the non-pecuniary rewards, as indicated by the fact that “piece control” generates an almost significant increase in pressure compared to “fixed control” (p-value = 0.1033).

We can accept H6. In the presence of pecuniary rewards, adding non-pecuniary rewards does not increase productivity or motivation, and does not decrease break time. This confirms the existence of a crowding-out effect, where any positive effect of gamification is removed when it is implemented at the same time as pecuniary rewards.

³³ To give some context, in both treatments 80% of the subjects did not take a single second of break, and if we remove the outlier, the maximum break time in “piece town” is 87 seconds.

Is gamification cost-effective?

To evaluate whether gamification could really be useful to a company, we can compare its cost-effectiveness to that of another intervention that yields the same performance level. The table below compares the performance and costs of the different treatments.

Table 41 - Performance and cost across the three main treatments

| Treatment | Average Performance (words found) | Cost per subject (€) |
|---------------|-----------------------------------|----------------------|
| Fixed control | 77 | 8 |
| Fixed badges | 86 | 8 |
| Fixed town | 91* | 8 |
| Piece control | 93* | 9.3* |
| Piece badges | 84 | 8.4 |
| Piece town | 87 | 8.7 |

*: significantly higher than "Fixed control" at the 10% significance level according to a Student test

This table clearly illustrates the main point of gamification: it can lead to the same increase in productivity as monetary incentives but without the equivalent increase in cost³⁴. Gamification, unlike piece rate, mainly consists on a fixed cost: the system has to be specifically designed to fit the tasks at hand, then it must be developed, and finally rolled-out. However, once in place, the maintenance cost is minimal, especially when compared to piece rate. This means that gamification will be way cheaper to scale up than piece rate, so in the long-run and in a large company, the difference in profitability will be sizable.

Part 6 - Discussion

There are several elements that this study teaches us. First, the way game design elements are implemented matters greatly. In fact, if we only tested pointification, as it is often the case (badges and/or leaderboards), we would not have found any significant positive impact on performance or intrinsic motivation. Indeed, it is only the "town building" treatment, which was our application of gamification, that was effective in our design. And it was effective because it made people achieve a higher performance level and let them feel more competent. The reason is that, in addition to providing pleasant visual rewards, this form of gamification gave the impression that they

³⁴ A Student test revealed no difference between "fixed town" and "piece control" in terms of performance, p-value = 0.9844. The difference in cost between "piece control" and "fixed town" is significant, with p-value = 0.08228).

were doing more than just solving puzzles: they were building something, and they could regularly see the results of their work getting better and better. Furthermore, this treatment gives a stronger impression of progression compared to the pointification treatment: in the latter, subjects earn badges one at a time, whereas they can earn 2 or 3 buildings at the same time, which makes them feel like they were accomplishing more, and thus were more competent.

In fact, the feeling of competence is the only element of intrinsic motivation that gamification improved. Despite what we expected, people did not find any implementation of non-pecuniary rewards to be more enjoyable than the control. This suggests that gamification being tied to the task may provide another intrinsic reason to perform well, but it does not make the task itself any more fun, even though it is the entire premise of the approach. If gamification had been enjoyable, it might have pushed players to take less break time, but it failed to do so. As we explained in the results section, we believe this is due to the task being more interesting than we expected. This left little room for improvement. However, we consider this to be a good surprise for gamification: it means that the approach can be effective even when it fails to provide more enjoyment, and that it goes beyond making tasks more fun... but only when it is not mixed with monetary incentives.

Indeed, another result is that none of the effects of gamification on performance held when we added monetary incentives. We found no difference in productivity or feeling of competence between the control group and groups that were exposed to both non-pecuniary and pecuniary rewards. It is important to remember that the effect on performance was driven by the feeling of competence, and that there was no significant difference in this feeling between “fixed control” and “fixed town”. We also know that there is one sizable difference between fixed payment and piece rate (besides the fact that one’s earnings are tied to one’s performance): in “fixed payment”, the gamification is the only form of performance feedback. Therefore, if a subject sees a lot of buildings appear, he will infer that his performance is quite high. However, under the piece rate scheme, the growth of earnings is another form of performance feedback. When exposed to the “town building” treatment and earning a fixed salary, seeing many new buildings appear might give the impression that one is performing well. However, when the fixed salary is replaced by a piece rate, noting that the earnings do not live up to the feeling of success provided by the gamification might moderate this enthusiasm, and diminish the feeling of competence.

There is another possible explanation for the lack of further performance increase when adding gamification alongside a piece rate scheme. The visual rewards may be a good way to assess one’s performance and flatter one’s ego in the absence of piece rate, but when the latter is

introduced, then it becomes the only thing that matters, and people just stop paying attention to the gamification. Therefore, decision makers should probably settle for one incentive, money or visual rewards, and not juxtapose different incentives. Mixing them is more costly, does not bring further benefit, and can even have adverse effects.

Indeed, the last surprising result we found was the increase in pressure felt in treatments that had both non-pecuniary rewards and pecuniary rewards. While we believed that non-pecuniary rewards would lead to a more pleasant experience, the opposite happened when combined with monetary incentives. However, as we saw later on, there is no significant difference in pressure between “piece control” and treatment that combined pecuniary and non-pecuniary rewards. This suggests that this increase in pressure was mainly driven by the monetary incentives. Nevertheless, this should be taken as a warning.

That being said, what are the main takeaways from this research? First, one should not mix too many forms of rewards. It is better to choose between monetary incentives and non-pecuniary rewards, rather than to implement both at the same time. So how should a decision-maker choose between one or the other? We have shown that gamification leads to the same increase in performance, while costing less than monetary incentives. For that reason alone, we think that trying gamification first is better, as the potential increase in profitability is higher, and if it does not work, it is always possible to replace it by monetary incentives. However, switching from monetary incentives to gamification alone may not go as smoothly. Obviously, employees will never appreciate a decrease in their earnings, and will severely dislike any gamification effort if they feel like it is the reason behind the decrease in their earnings. Another more insidious reason why this is a bad idea is the crowding-out effect: if monetary incentives have been in place for a while, they will have severely undermined the intrinsic motivation for the task. In our design, gamification improved the feeling of competence in the absence of monetary incentives, but we did not test what would happen if we had a piece rate for a part of the experiment, and only the gamification for the rest of the experiment. For these two reasons, we believe that testing monetary incentives before gamification is very risky: if they fail to increase productivity enough to justify their cost, it will be very difficult to go back without suffering from a severe backlash.

In the end, if one has not implemented monetary incentives yet, we recommend testing out gamification, as it is a cheaper and safer alternative way to improve both productivity and intrinsic motivation.

Limitations and further research

There are of course limitations to this study. First, the experiment was short, and one might wonder whether our results would hold if subjects worked for several hours rather than just 30 minutes. Nonetheless, the results that we found are promising. Besides, as the difference in performance between treatments only got bigger as time went by, our results suggest that this effect might be stronger if the experiment lasted longer. This is why it would be worth replicating this experiment on a much longer timeframe, using something like Amazon Mechanical Turk for example.

Second, we needed to have a simple task with a clear quantifiable outcome, but a lot of jobs nowadays involve activities that are way more complex than this. This is why, unfortunately, the results of our study cannot be generalized to more complex activities. However, there are still a lot of activities for which these results can be applied. We can for instance think about the employees of a call center, manual workers, salesmen ... All of these activities fit the criteria that we define, and yet their employers rarely rely go beyond pointification to motivate their workforce. Besides, more complex tasks mean that the gamification system used to support them would have to be more elaborate. And, as those complex works tend to generate more added values, gamification does show a lot of potential. This is why we think that such an experiment should be replicated with a more complicated task.

In addition to overcoming the limitations of our study, there are other ways to conduct further research. First, our data may suggest that the gamification was more effective on people who were already good at the task to begin with. Unfortunately, our sample was too small for us to draw definitive conclusions on this topic, and one could replicate our study with a much larger sample, or directly on the field rather than in a laboratory.

Second, nudges are often criticized for their lack of long-run effect, and the same criticism might arise about what we propose. We believe that if the gamification is regularly renewed (for instance, by changing the pictures every month), the effect could be sustainable in the long-run, but this remains to be tested. Finally, if the effect of adding gamification is comparable to the effect of adding a piece rate, then it is entirely possible that a crowding-out effect could occur. For example, what would happen if gamification was added to a work environment where employees are highly driven? We believe that this would be a very interesting question to answer.

Applications

There are several ways to apply our findings. First, gamification should be seriously considered by managers whose teams work on tasks that are comparable to the one in our study. Generally, any type of activity in which there is a clear, quantifiable outcome, should be a good candidate for gamification. A typical type of activity that comes to mind is sales, where monetary incentives and leaderboards are already implemented. Not only can gamification be as effective as monetary incentives for improving performance, but it is definitely better than these incentives for improving employee intrinsic motivation, and as such should be part of every manager's toolkit.

And for people who already started to implement gamification but only with leaderboards and badges, here is the second application of this study: one should not rely exclusively on pointification. We demonstrated here that going beyond simple badges and medals can completely change the results of the intervention, and so we urge people who already implemented leaderboards and badges and don't intend to take their initiative any further to reconsider their position. Of course, these elements can work notably because it leads to a higher social status or because they are tied to monetary rewards or promotions, and there is no such thing in our design. However, the "town building" treatment is not incompatible with social proof elements, on the contrary, so there is no real reason not to use the better of the two forms of non-pecuniary rewards.

Before concluding, we want to clarify that we certainly do not mean that extrinsic incentives are strictly inferior to gamification. We already showed that there was a lot of documented evidence of their effectiveness. We merely want to point out that they are not without problems, and that they are certainly not the best and only way to improve motivation. The real aim of our study is to encourage further investigations and a deeper reflection on this topic. It is to show how gamification techniques could be more suitable for particular contexts and working tasks in order to enhance long term engagement and sustainability in performance. With this study, we hope to shed a new light on gamification's potential to improve intrinsic motivation when engagement at work is lacking.

Conclusion

In this chapter, we introduced the problem of motivation at work, and proposed to apply the Nudge approach via a particular prism: gamification. We then built an economic model of gamification at work to show how it worked and how it interacted with monetary rewards. Then, we tested gamification and compared it to monetary rewards and to pointification, a form of gamification which mostly relies on badges and leaderboards. We found that gamification was effective at increasing productivity and intrinsic motivation. We ended this paper by discussing our findings, presenting the limits of this study and proposing directions for future researches.

While we hardly ever employed the word in this chapter, it is very much an application of Nudging: we improved productivity simply by reframing the task, which was no longer just about finding words, but about building a virtual town or earning badges. A simple change which resulted in a significant difference. We hope this paper allowed managers to find out about new ways to improve the motivation of their employees in a way that will make everybody better off.

Conclusion

In this thesis, we started by presenting the issues that the classical economic approach struggled with, before introducing the Nudge approach and showing its interest for solving them. Then, we applied this approach to the issue of web survey participation. We found that it significantly increased participation, but only when implemented at the right time and in the right way. This allowed us to highlight the importance of context and provide some insights on the drivers of participation. After that, we used the underlying science behind the Nudge approach to develop an alternative way of measuring job satisfaction, and we tested it on interns. We found that it allowed us to collect relevant information at an earlier date and in a more convenient way than the classical measure. This will give the managers and human resources the tools they need to detect disappointed employees as early as possible. In our last chapter, we used the Nudge approach to make a repetitive task more enjoyable, with the hope that it would increase intrinsic motivation and productivity. By directly comparing this approach to monetary incentives, we found out that Nudging could generate the same increase in productivity as financial rewards, while being way less costly in the long run.

Of course, these were just a few examples of Nudges. Nevertheless, we hope they provided useful insights and showed that Nudges can be successfully applied in a great variety of contexts to help business firms, but also organizations and governments. Indeed, this work goes beyond just showing that Nudging can be used to improve survey participation, satisfaction measures and productivity. We wanted to illustrate the fact that, using this approach, it is possible to solve practically any problem, even those beyond the scope of economic science. No matter the issue, the general method remains the same: get a good understanding of the problem (via literature review, or simply by collecting insights and testimonials from people working in the environment you want to Nudge), keep monitoring the behavioral economics and Nudge literature to see what creative solutions others have come up with, and then think about which of those Nudges would be a good fit for the current problem. If nothing can be directly applied (for instance if you are interested in solving a problem for which Nudging was not used yet), then it is just a matter of paying very good attention to details. In Nudging, perhaps more than in anything else, the devil is in the details. It is a few tweaks in wording that made some of the Nudges successful at improving participation where others failed, after all. Then, all that is left to do is to test, learn and adapt.

With that in mind, we shall conclude this work by encouraging the reader to try the Nudge approach to solve one of their current issues. We hope that this work was useful and interesting, and we wish you a happy nudging!

Appendix

Happy Trainees questionnaires

Aspirations

All « aspiration » items started by “I want to [...]”

Progression

1. Put in practice what I studied
2. Get progressively integrated into the professional world
3. Deepen one or several expertises

Environment

1. Feel useful quickly
2. Have a good work/life balance
3. Join a stimulating team

Management

1. Feel like my work is being recognized
2. Work with an inspiring manager
3. Have clear objectives for my internship

Motivation

1. Face intellectual challenges
2. Be granted real responsibilities as an intern
3. Be autonomous

Pride

1. Realize my internship in the company I want to be hired by
2. Accomplish something I can be proud of
3. Have a competitive salary

Fun

1. Perform an activity in relation to my interests and values
2. Be 100% committed to a project
3. Join a real community of interns

Evaluations

Progression

1. I felt like I acquired useful skills
2. I felt like I have grown professionally
3. I could benefit from the experience of competent teams

Environnement

1. I was quickly integrated to the daily activities of my team

2. I was satisfied with my work/life balance
3. I think that my company is innovative

Management

1. I felt encouraged in my efforts
2. I could benefit from regular and constructive feedback from my manager
3. I understood what was expected of me well

Motivation

1. I felt committed to my mission
2. I felt like I was given appropriate means to perform my mission
3. I felt like I was being trusted

Fierté

1. I would like to be hired by this company
2. I recommend this company for an internship.
3. I felt like my salary was competitive compared to what I could have gotten in other companies

Fun

1. I found meaning in what I did.
2. I lived an intense experience.
3. I enjoyed the quality of human relationships (friendliness, conviviality, ...) in my company.

Appendix 2 – Correlation tables between aspirations, evaluations, final satisfaction and internship grade

Table of correlations between evaluations and final satisfaction (N = 101)

| | eval_envir | eval_pride | eval_motivation | eval_fun | eval_management | eval_progres |
|--------------------|------------|------------|-----------------|----------|-----------------|--------------|
| eval_envir | | | | | | |
| eval_pride | 0.57**** | | | | | |
| eval_motivation | 0.49**** | 0.55**** | | | | |
| eval_fun | 0.60**** | 0.67**** | 0.62**** | | | |
| eval_management | 0.60**** | 0.59**** | 0.69**** | 0.61**** | | |
| eval_progres | 0.42**** | 0.59**** | 0.64**** | 0.69**** | 0.68**** | |
| final_satisfaction | 0.49**** | 0.56**** | 0.60**** | 0.60**** | 0.60**** | 0.67**** |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table of correlations between aspirations and final satisfaction (N = 101)

| | aspi_envir | aspi_pride | aspi_motivation | aspi_fun | aspi_management | aspi_progres |
|--------------------|------------|------------|-----------------|----------|-----------------|--------------|
| aspi_envir | | | | | | |
| aspi_pride | 0.23* | | | | | |
| aspi_motivation | 0.17 | 0.36*** | | | | |
| aspi_fun | 0.24* | 0.47**** | 0.40**** | | | |
| aspi_management | 0.42**** | 0.53**** | 0.40**** | 0.37*** | | |
| aspi_progres | 0.14 | 0.40**** | 0.34*** | 0.48**** | 0.29** | |
| final_satisfaction | 0.11 | -0.09 | 0.20* | -0.02 | -0.03 | 0.04 |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table of correlations between evaluations, final satisfaction and grade (N = 64)

| | eval_envir | eval_pride | eval_motivation | eval_fun | eval_management | eval_progres | final_satisfaction |
|--------------------|------------|------------|-----------------|----------|-----------------|--------------|--------------------|
| eval_envir | | | | | | | |
| eval_pride | 0.60**** | | | | | | |
| eval_motivation | 0.48**** | 0.53**** | | | | | |
| eval_fun | 0.59**** | 0.68**** | 0.59**** | | | | |
| eval_management | 0.54**** | 0.60**** | 0.66**** | 0.67**** | | | |
| eval_progres | 0.25* | 0.60**** | 0.54**** | 0.71**** | 0.64**** | | |
| final_satisfaction | 0.43*** | 0.60**** | 0.66**** | 0.74**** | 0.65**** | 0.67**** | |
| internship_grade | 0.01 | 0.25* | 0.27* | 0.13 | 0.23 | 0.14 | 0.29* |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table of correlations between aspirations, final satisfaction and grade (N = 64)

| | aspi_envir | aspi_pride | aspi_motivation | aspi_fun | aspi_management | aspi_progres | final_satisfaction |
|--------------------|------------|------------|-----------------|----------|-----------------|--------------|--------------------|
| aspi_envir | | | | | | | |
| aspi_pride | 0.30* | | | | | | |
| aspi_motivation | 0.16 | 0.37** | | | | | |
| aspi_fun | 0.24 | 0.39** | 0.44*** | | | | |
| aspi_management | 0.48**** | 0.53**** | 0.41*** | 0.26* | | | |
| aspi_progres | 0.02 | 0.30* | 0.36** | 0.40** | 0.26* | | |
| final_satisfaction | -0.04 | 0.04 | 0.22 | 0.20 | 0.07 | 0.17 | |
| internship_grade | 0.18 | 0.06 | 0.13 | -0.05 | 0.27* | -0.14 | 0.29* |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table of correlations between “progression” items (N = 101)

| | eval_progres_1 | eval_progres_2 | eval_progres_3 | aspi_progres_1 | aspi_progres_2 |
|----------------|----------------|----------------|----------------|----------------|----------------|
| eval_progres_1 | | | | | |
| eval_progres_2 | 0.65**** | | | | |
| eval_progres_3 | 0.43**** | 0.36*** | | | |
| aspi_progres_1 | 0.07 | 0.14 | -0.06 | | |
| aspi_progres_2 | 0.03 | 0.03 | 0.08 | -0.03 | |
| aspi_progres_3 | -0.04 | 0.11 | 0.01 | 0.27** | 0.30** |

p < .0001 ‘****’; p < .001 ‘***’, p < .01 ‘**’, p < .05 ‘*’

Table of correlations between “management” items (N = 101)

| | eval_management_1 | eval_management_2 | eval_management_3 | aspi_management_1 | aspi_management_2 |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| eval_management_1 | | | | | |
| eval_management_2 | 0.54**** | | | | |
| eval_management_3 | 0.32** | 0.52**** | | | |
| aspi_management_1 | -0.06 | 0.04 | 0.23* | | |
| aspi_management_2 | 0.11 | 0.13 | 0.19 | 0.27** | |
| aspi_management_3 | 0.00 | 0.03 | -0.01 | 0.19 | 0.40**** |

p < .0001 ‘****’; p < .001 ‘***’, p < .01 ‘**’, p < .05 ‘*’

Table of correlations between “fun” items (N = 101)

| | eval_fun_1 | eval_fun_2 | eval_fun_3 | aspi_fun_1 | aspi_fun_2 |
|------------|------------|------------|------------|------------|------------|
| eval_fun_1 | | | | | |
| eval_fun_2 | 0.45**** | | | | |
| eval_fun_3 | 0.35**** | 0.32*** | | | |
| aspi_fun_1 | 0.16 | 0.28** | 0.13 | | |
| aspi_fun_2 | 0.25* | 0.29** | 0.00 | 0.09 | |
| aspi_fun_3 | -0.08 | 0.03 | -0.03 | -0.06 | 0.06 |

p < .0001 ‘****’; p < .001 ‘***’, p < .01 ‘**’, p < .05 ‘*’

Table of correlations between “environment” items (N = 101)

| | eval_envir_1 | eval_envir_2 | eval_envir_3 | aspi_envir_1 | aspi_envir_2 |
|--------------|--------------|--------------|--------------|--------------|--------------|
| eval_envir_1 | | | | | |
| eval_envir_2 | 0.34*** | | | | |
| eval_envir_3 | 0.31** | 0.04 | | | |
| aspi_envir_1 | 0.19 | 0.04 | 0.13 | | |
| aspi_envir_2 | -0.05 | 0.01 | 0.06 | 0.14 | |
| aspi_envir_3 | 0.28** | -0.09 | 0.07 | 0.24* | 0.04 |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table of correlations between “motivation” items (N = 101)

| | eval_motivation_1 | eval_motivation_2 | eval_motivation_3 | aspi_motivation_1 | aspi_motivation_2 |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| eval_motivation_1 | | | | | |
| eval_motivation_2 | 0.41**** | | | | |
| eval_motivation_3 | 0.62**** | 0.39**** | | | |
| aspi_motivation_1 | 0.12 | 0.08 | 0.10 | | |
| aspi_motivation_2 | 0.04 | 0.02 | 0.06 | 0.24* | |
| aspi_motivation_3 | 0.07 | 0.13 | 0.10 | 0.12 | 0.40**** |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table of correlations between “pride” items (N = 101)

| | eval_pride_1 | eval_pride_2 | eval_pride_3 | aspi_pride_1 | aspi_pride_2 |
|--------------|--------------|--------------|--------------|--------------|--------------|
| eval_pride_1 | | | | | |
| eval_pride_2 | 0.55**** | | | | |
| eval_pride_3 | 0.19 | 0.46**** | | | |
| aspi_pride_1 | 0.19 | 0.07 | -0.01 | | |
| aspi_pride_2 | 0.10 | 0.00 | -0.03 | 0.29** | |
| aspi_pride_3 | 0.02 | -0.15 | 0.00 | 0.23* | 0.26** |

p < .0001 '****'; p < .001 '***', p < .01 '**', p < .05 '*'

Table of correlation between “evaluation” dimensions

| | | | | | | |
|--------------------|-------------------|-------------------|------------------------|-----------------|------------------------|---------------------|
| | eval_envir | eval_pride | eval_motivation | eval_fun | eval_management | eval_progres |
| eval_envir | | | | | | |
| eval_pride | 0.57**** | | | | | |
| eval_motivation | 0.49**** | 0.55**** | | | | |
| eval_fun | 0.60**** | 0.67**** | 0.62**** | | | |
| eval_management | 0.60**** | 0.59**** | 0.69**** | 0.61**** | | |
| eval_progres | 0.42**** | 0.59**** | 0.64**** | 0.69**** | 0.68**** | |
| final_satisfaction | 0.49**** | 0.56**** | 0.60**** | 0.60**** | 0.60**** | 0.67**** |

Table of correlation between “aspiration” dimensions

| | | | | | | |
|--------------------|-------------------|-------------------|------------------------|-----------------|------------------------|---------------------|
| | aspi_envir | aspi_pride | aspi_motivation | aspi_fun | aspi_management | aspi_progres |
| aspi_envir | | | | | | |
| aspi_pride | 0.23* | | | | | |
| aspi_motivation | 0.17 | 0.36*** | | | | |
| aspi_fun | 0.24* | 0.47**** | 0.40**** | | | |
| aspi_management | 0.42**** | 0.53**** | 0.40**** | 0.37*** | | |
| aspi_progres | 0.14 | 0.40**** | 0.34*** | 0.48**** | 0.29** | |
| final_satisfaction | 0.11 | -0.09 | 0.20* | -0.02 | -0.03 | 0.04 |

Appendix 3 – Intermediate regression of internship grade and final satisfaction on various items of evaluation and aspirations

Regression of final satisfaction on all three items of “progress” evaluation

| <i>Predictors</i> | final satisfaction | | | final satisfaction | | | final satisfaction | | |
|--|---------------------------|-------------|------------------|---------------------------|-------------|------------------|---------------------------|-------------|------------------|
| | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 1.96 | 0.22 – 3.70 | 0.029 | 2.32 | 0.70 – 3.94 | 0.006 | 1.77 | 0.15 – 3.40 | 0.035 |
| extraversion | 0.14 | 0.04 – 0.24 | 0.007 | 0.12 | 0.02 – 0.22 | 0.024 | 0.11 | 0.01 – 0.21 | 0.028 |
| eval progres 1 | 1.02 | 0.68 – 1.37 | <0.001 | | | | | | |
| eval progres 2 | | | | 0.98 | 0.66 – 1.31 | <0.001 | | | |
| eval progres 3 | | | | | | | 1.12 | 0.79 – 1.46 | <0.001 |
| Observations | 101 | | | 101 | | | 101 | | |
| R ² / adjusted R ² | 0.316 / 0.302 | | | 0.325 / 0.311 | | | 0.365 / 0.352 | | |

Regression of internship grade on all three items of “management” aspiration

| <i>Predictors</i> | internship grade | | | internship grade | | | internship grade | | |
|--|------------------|--------------|--------------|------------------|--------------|----------|------------------|--------------|--------------|
| | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 11.73 | 8.82 – 14.64 | <0.001 | 12.46 | 9.88 – 15.04 | <0.001 | 12.41 | 9.65 – 15.16 | <0.001 |
| eval motivation | 0.64 | 0.08 – 1.20 | 0.028 | 0.57 | -0.00 – 1.14 | 0.056 | 0.62 | 0.06 – 1.19 | 0.035 |
| aspi management 1 | 0.38 | -0.05 – 0.82 | 0.090 | | | | | | |
| aspi management 2 | | | | 0.27 | -0.07 – 0.61 | 0.119 | | | |
| aspi management 3 | | | | | | | 0.22 | -0.15 – 0.59 | 0.251 |
| Observations | 64 | | | 64 | | | 64 | | |
| R ² / adjusted R ² | 0.115 / 0.086 | | | 0.109 / 0.079 | | | 0.092 / 0.062 | | |

Regression of internship grade on all three items of “motivation” evaluation

| <i>Predictors</i> | internship grade | | | internship grade | | | internship grade | | |
|--|------------------|--------------|--------------|------------------|---------------|----------|------------------|--------------|--------------|
| | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> | <i>Estimates</i> | <i>CI</i> | <i>p</i> |
| (Intercept) | 11.67 | 8.90 – 14.44 | <0.001 | 12.87 | 10.44 – 15.29 | <0.001 | 12.04 | 9.37 – 14.71 | <0.001 |
| aspi management | 0.52 | 0.01 – 1.03 | 0.049 | 0.53 | 0.01 – 1.06 | 0.052 | 0.60 | 0.09 – 1.11 | 0.025 |
| eval motivation 1 | 0.54 | 0.02 – 1.06 | 0.048 | | | | | | |
| eval motivation 2 | | | | 0.27 | -0.17 – 0.71 | 0.238 | | | |
| eval motivation 3 | | | | | | | 0.38 | -0.04 – 0.80 | 0.078 |
| Observations | 64 | | | 64 | | | 64 | | |
| R ² / adjusted R ² | 0.130 / 0.102 | | | 0.093 / 0.063 | | | 0.118 / 0.089 | | |

References

- Akerlof, G. A., & Yellen, J. L. (1985). A near-rational model of the business cycle, with wage and price inertia. *The Quarterly Journal of Economics*, 100(Supplement), 823-838.
- Allais, M. (1953). Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. *Econometrica*, 21(4), 503. <https://doi.org/10.2307/1907921>
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9), 1082-1095.
- Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10), 3003-37.
- Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response Rates in Organizational Science, 1995–2008: A Meta-analytic Review and Guidelines for Survey Researchers. *Journal of Business and Psychology*, 25(3), 335-349. <https://doi.org/10.1007/s10869-010-9157-6>
- Ariely, Dan (2008). *Predictably Irrational*, Harper Collins. ISBN 978-0-06-135323-9
- Ashkanasy, N. M., Härtel, C. E., & Zerbe, W. J. (2000). *Emotions in the workplace: Research, theory, and practice*. Quorum Books/Greenwood Publishing Group.
- Avey, J. B., Reichard, R. J., Luthans, F., & Mhatre, K. H. (2011). Meta-analysis of the impact of positive psychological capital on employee attitudes, behaviors, and performance. *Human Resource Development Quarterly*, 22(2), 127-152. <https://doi.org/10.1002/hrdq.20070>
- Avineri, E. (2012). On the use and potential of behavioural economics from the perspective of transport and climate change. *Journal of Transport Geography*, 24, 512-521.
- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139-1160. <https://doi.org/10.1177/0018726708094863>
- Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., ... & Galing, S. (2017). Should governments invest more in nudging?. *Psychological science*, 28(8), 1041-1055.
- Bono, J. E., Foldes, H. J., Vinson, G., & Muros, J. P. (2007). Workplace emotions: The role of supervision and leadership. *Journal of Applied Psychology*, 92(5), 1357.
- Bosnjak, M., Tuten, T. L., & Wittmann, W. W. (2005). Unit (non)response in Web-based access panel surveys: An extended planned-behavior approach. *Psychology and Marketing*, 22(6), 489-505. <https://doi.org/10.1002/mar.20070>
- Brandon, A., Ferraro, P. J., List, J. A., Metcalfe, R. D., Price, M. K., & Rundhammer, F. (2017). Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments (No. w23277). National Bureau of Economic Research.
- Brick, J. M., & Williams, D. (2013). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 36-59. <https://doi.org/10.1177/0002716212456834>

- Brown, C. L., & Krishna, A. (2004). The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice. *Journal of Consumer Research*, 31(3), 529-539. <https://doi.org/10.1086/425087>
- Bruce, S., Hawkins, P., Sharp, M., & Keller, A. (2006). Experienced-based suggestions for achieving a high survey response rate. Retrieved from <http://www.virginia.edu/case/education/documents/surveyresponseratesummarypaper-final.pdf>.
- Caillois, R. (1961). *Man, Play, and Games*. New York.
- Camerer, C. F., Loewenstein, G., & Rabin, M. (2011). *Advances in Behavioral Economics*. Princeton University Press.
- Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Regulation for Conservatives: Behavioral Economics and the Case for « Asymmetric Paternalism ». *University of Pennsylvania Law Review*, 151(3), 1211. <https://doi.org/10.2307/3312889>
- Cammann, C., Fichman, M., Jenkins, D., & Klesh, J. (1979). The Michigan organizational assessment questionnaire. *Unpublished manuscript, University of Michigan, Ann Arbor*.
- Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Optimal defaults and active decisions. *The Quarterly Journal of Economics*, 124(4), 1639-1674.
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology* (Vol. 24, pp. 201-234). Academic Press.
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6), 1015-1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- Clark, A. E., & Georgellis, Y. (2004). *Kahneman meets the quitters: Peak-end behaviour in the labour market. CNRS and DELTA*.
- Coco, M. (2000). Internships: A try before you buy arrangement. *SAM Advanced Management Journal*, 65(2), 41.
- Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, 60(6), 821-836.
- Couper, M. P., & Miller, P. V. (2008). Web Survey Methods: Introduction. *Public Opinion Quarterly*, 72(5), 831-835. <https://doi.org/10.1093/poq/nfn066>
- Cronqvist, H., Thaler, R. H., & Yu, F. (2018). When Nudges Are Forever: Inertia in the Swedish Premium Pension Plan. In *AEA Papers and Proceedings* (Vol. 108, pp. 153-58).
- Cronqvist, H., Thaler, R. H., & Yu, F. (2018). When Nudges Are Forever: Inertia in the Swedish Premium Pension Plan. In *AEA Papers and Proceedings* (Vol. 108, pp. 153-58).
- Curtin, R., Presser, S., & Singer, E. (2005). Changes in Telephone Survey Nonresponse over the Past Quarter Century. *Public Opinion Quarterly*, 69(1), 87-98. <https://doi.org/10.1093/poq/nfi002>
- Daugherty, S. R., DeWitt C. Baldwin, J., & Rowley, B. D. (1998). Learning, Satisfaction, and Mistreatment During Medical Internship: A National Survey of Working Conditions. *JAMA*, 279(15), 1194-1199. <https://doi.org/10.1001/jama.279.15.1194>

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668. <https://doi.org/10.1037/0033-2909.125.6.627>

Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From game design elements to gamefulness: Defining « gamification »*. 9. <https://doi.org/10.1145/2181037.2181040>

Di Lorenzo-Aiss, J., & Mathisen, R. E. (1996). Marketing higher education: Models of marketing internship programs as tools for the recruitment and retention of undergraduate marketing majors. *Journal of marketing for higher education*, 7(1), 71–84.

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in Education: A Systematic Mapping Study. *Journal of Educational Technology & Society*, 18(3), 75-88.

Divine, R. L., Linrud, J. K., Miller, R. H., & Wilson, J. H. (2015). Required Internship Programs in Marketing: Benefits, Challenges and Determinants of Fit. *Marketing Education Review*. Consulté à l'adresse <https://www.tandfonline.com/doi/abs/10.1080/10528008.2007.11489003>

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Ft Worth, TX, US: Dryden Press.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 76, 643-669.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26(2), 132-139. <https://doi.org/10.1016/j.chb.2009.10.015>

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114, 817–868.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, 58(2), 203-210. <http://dx.doi.org/10.1037/h0041593>

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, 65(1), 45.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: the foot-in-the-door technique. *Journal of personality and social psychology*, 4(2), 195.

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *The Journal of Political Economy*, 279–304.

Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. In *Survey Research Methods* (Vol. 2, p. 21–32). Consulté à l'adresse https://www.researchgate.net/profile/Stephane_Ganassali/publication/264138574_The_Influence_of_the_Design_of_Web_Survey_Questionnaires_on_the_Quality_of_Responses/links/53cfab290cf25dc05cfb0c47.pdf

Gault, J., Redington, J., & Schlager, T. (2000). Undergraduate business internships and career success: Are they related? *Journal of marketing education*, 22(1), 45–53.

- Georgellis, Y., Iossa, E., & Tabvuma, V. (2010). Crowding out intrinsic motivation in the public sector. *Journal of Public Administration Research and Theory*, 21(3), 473-493.
- Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-015-0248-1>
- Gkorezis, P., & Kastritsi, A. (2017). Employee expectations and intrinsic motivation: Work-related boredom as a mediator. *Employee Relations*, 39(1), 100-111. <https://doi.org/10.1108/ER-02-2016-0025>
- Goodwin, T. (2012). Why We Should Reject 'Nudge'. *Politics*, 32(2), 85-92. <https://doi.org/10.1111/j.1467-9256.2012.01430.x>
- Goritz, A. S. (2008). The Long-Term Effect of Material Incentives on Participation in Online Panels. *Field Methods*, 20(3), 211-225. <https://doi.org/10.1177/1525822X08317069>
- Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504-528.
- Grant, A. M., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of Personality and Social Psychology*, 98(6), 946-955. <https://doi.org/10.1037/a0017935>
- Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.
- Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56(4), 475-495.
- Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: description and an illustration. *The Public Opinion Quarterly*, 64(3), 299-308.
- Grüne-Yanoff, T. (2012). Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, 38(4), 635-645. <https://doi.org/10.1007/s00355-011-0636-0>
- Gupta, P., Burns, D. J., & Schiferl, J. S. (2010). *An Exploration of Student Satisfaction with Internship Experiences in Marketing*. Consulté à l'adresse <https://papers.ssrn.com/abstract=1871152>
- Handgraaf, M. J., de Jeude, M. A. V. L., & Appelt, K. C. (2013). Public praise vs. private pay: Effects of rewards on energy conservation in the workplace. *Ecological Economics*, 86, 86-92.
- Hansen, P. G. (2015). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation*, (1), 1-20.
- Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. *European Journal of Risk Regulation*, 2013, 3.
- Hassett, K. A., & Hubbard, R. G. (2002). Tax policy and business investment. In *Handbook of public economics* (Vol. 3, pp. 1293-1343). Elsevier.
- Hausman, D. M., & Welch, B. (2010). Debate: To Nudge or Not to Nudge*. *Journal of Political Philosophy*, 18(1), 123-136. <https://doi.org/10.1111/j.1467-9760.2009.00351.x>

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

Heerwegh, D. (2006). An Investigation of the Effect of Lotteries on Web Survey Response Rates. *Field Methods*, 18(2), 205-220. <https://doi.org/10.1177/1525822X05285781>

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method : Measuring the quality of everyday life*. Sage.

Holland, R. W., Aarts, H., & Langendam, D. (2006). Breaking and creating habits on the working floor: A field-experiment on the power of implementation intentions. *Journal of Experimental Social Psychology*, 42(6), 776-783.

Hong, K.-S. (2002). Relationships between students' and instructional variables with satisfaction and learning from a Web-based course. *The Internet and Higher Education*, 5(3), 267–281.

Hossain, T., & List, J. A. (2012). The behaviorist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, 58(12), 2151-2167.

Huotari, K., & Hamari, J. (2012). Defining gamification: A service marketing perspective. *Proceeding of the 16th International Academic MindTrek Conference*, 17–22. Consulté à l'adresse <http://dl.acm.org/citation.cfm?id=2393137>

Jenkins Jr, G. D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance? A meta-analytic review of empirical research. *Journal of applied psychology*, 83(5), 777.

Johnson, E. J., & Goldstein, D. G. (2003). Do defaults save lives? *Science*, 302, 1338–1339.

Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of interactive learning research*, 11(1), 29.

Joinson, A. N., & Reips, U.-D. (2007). Personalized salutation, power of sender and response rates to Web-based surveys. *Computers in Human Behavior*, 23(3), 1372-1383. <https://doi.org/10.1016/j.chb.2004.12.011>

Kahneman, D. (2000). Evaluation by moments : Past and future. *Choices, values, and frames*, 693–708.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263- 291. <https://doi.org/10.2307/1914185>

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The quarterly journal of economics*, 112(2), 375–406.

Kaplowitz, M. D., Lupi, F., Couper, M. P., & Thorp, L. (2012). The Effect of Invitation Design on Web Survey Response Rates. *Social Science Computer Review*, 30(3), 339-349. <https://doi.org/10.1177/0894439311419084>

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.

Kifetew, F. M., Munante, D., Perini, A., Susi, A., Siena, A., Busetta, P., & Valerio, D. (2017). Gamifying Collaborative Prioritization: Does Pointsification Work? *2017 IEEE 25th*

International Requirements Engineering Conference (RE), 322-331.
<https://doi.org/10.1109/RE.2017.66>

King, D., Greaves, F., Exeter, C., & Darzi, A. (2013). *'Gamification': Influencing health behaviours with games*. SAGE Publications Sage UK: London, England.

King, N. (1970). Clarification and evaluation of the two-factor theory of job satisfaction. *Psychological Bulletin*, 74(1), 18.

Klor, E. F., Kube, S., Winter, E., & Zultan, R. (2014). Can higher rewards lead to less effort? Incentive reversal in teams. *Journal of Economic Behavior & Organization*, 97, 72-83.
<https://doi.org/10.1016/j.jebo.2013.10.010>

Kosfeld, M., & Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3), 86-99.

Lee, S. J., Srinivasan, S., Trail, T., Lewis, D., & Lopez, S. (2011). Examining the relationship among student perception of support, course satisfaction, and learning outcomes in online learning. *The Internet and Higher Education*, 14(3), 158-163.

Lévy-Garboua, L., Montmarquette, C., & Simonnet, V. (2001). *Job satisfaction and quits : Theory and evidence from the German Socioeconomic Panel*. CIRANO.

Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a fad? Gamification in health and fitness apps. *JMIR serious games*, 2(2).

Liu, P. J., Wisdom, J., Roberto, C. A., Liu, L. J., & Ubel, P. A. (2014). Using Behavioral Economics to Design More Effective Food Policies to Address Obesity. *Applied Economic Perspectives and Policy*, 36(1), 6-24. <https://doi.org/10.1093/aep/ppt027>

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267-286. <https://doi.org/10.1037/0033-2909.127.2.267>

Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2014). Warning: You are About to Be Nudged. Available at SSRN 2417383. Consulté à l'adresse http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2417383

Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly journal of economics*, 116(4), 1149-1187.

Maidani, E. A. (1991). Comparative study of Herzberg's two-factor theory of job satisfaction among public and private sectors. *Public personnel management*, 20(4), 441-448.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., Vehovar, V., & Berzelak, N. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *Journal of the Market Research Society*, 50(1), 79.

Markowitz, H. (1952). The utility of wealth. *The Journal of Political Economy*, 151-158.

Massung, E., Coyle, D., Cater, K. F., Jay, M., & Preist, C. (2013). Using crowdsourcing to support pro-environmental community activism. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 371.
<https://doi.org/10.1145/2470654.2470708>

Matallaoui, A., Hanner, N., & Zarnekow, R. (2017). Introduction to Gamification: Foundation and Underlying Theories. In S. Stieglitz, C. Lattemann, S. Robra-Bissantz, R. Zarnekow, & T.

- Brockmann (Éd.), *Gamification: Using Game Elements in Serious Contexts* (p. 3-18). https://doi.org/10.1007/978-3-319-45557-0_1
- McAllister, D. J., & Bigley, G. A. (2002). Work Context and the Definition of self: How Organizational Care Influences Organization-Based Self-Esteem. *Academy of Management Journal*, 45(5), 894-904. <https://doi.org/10.5465/3069320>
- McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.
- Mullainathan, S., & Thaler, R. H. (2000). Behavioral Economics (Working Paper N° 7948). Consulté à l'adresse National Bureau of Economic Research website: <http://www.nber.org/papers/w7948>
- Nevid, J. (2012). *Psychology: Concepts and applications*. Nelson Education.
- O'Boyle, E., & Harter, J. (2017). State of the American workplace report.
- O'Boyle, E., & Harter, J. (2017). *State of the American workplace report*.
- Pankiewicz, M. (2016). DATA ANALYSIS FOR MEASURING EFFECTS OF GAMIFICATION IN E-LEARNING ENVIRONMENTS. *EDULEARN16 Proceedings*, 7082-7088.
- Paulins, V. A. (2008). Characteristics of retailing internships contributing to students' reported satisfaction with career development. *Journal of Fashion Marketing and Management: An International Journal*, 12(1), 105–118.
- Pedersen, M. J., & Nielsen, C. V. (2014). Improving Survey Response Rates in Online Panels: Effects of Low-Cost Incentives and Cost-Free Text Appeal Interventions. *Social Science Computer Review*. <https://doi.org/10.1177/0894439314563916>
- Pichert, D., & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behaviour. *Journal of Environmental Psychology*, 28(1), 63-73. <https://doi.org/10.1016/j.jenvp.2007.09.004>
- Pierce, J. L., & Gardner, D. G. (2004). Self-Esteem Within the Work and Organizational Context: A Review of the Organization-Based Self-Esteem Literature. *Journal of Management*, 30(5), 591-622. <https://doi.org/10.1016/j.jm.2003.10.001>
- Porter, S. R. (2005). E-mail Subject Lines and Their Effect on Web Survey Viewing and Response. *Social Science Computer Review*, 23(3), 380-387. <https://doi.org/10.1177/0894439305275912>
- Porter, S. R., & Whitcomb, M. E. (2003). The impact of contact type on web survey response rates. *The Public Opinion Quarterly*, 67(4), 579–588.
- Rabin, M. (1998). Psychology and economics. *Journal of economic literature*, 36(1), 11-46.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353.
- Romaniuc, R. (2017). Intrinsic motivation in economics: A history. *Journal of Behavioral and Experimental Economics*, 67, 56–64.
- Sauer, S. J., Rodgers, M., & Becker, W. J. (2018). The Effects of Goals and Pay Structure on Managerial Reporting Dishonesty. *Journal of Accounting, Ethics and Public Policy*, 19(3).

Schoeni, R. F., Stafford, F., Mcgonagle, K. A., & Andreski, P. (2013). Response Rates in National Panel Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 60-87. <https://doi.org/10.1177/0002716212456363>

Schreiber, C. A., & Kahneman, D. (1996). *Beyond the peak and end hypothesis : Exploring the relation between real-time displeasure and retrospective evaluation*. Working Paper, Princeton University.

Scitovsky, T. (1976). *The joyless economy: An inquiry into human satisfaction and consumer dissatisfaction*. Oxford, England: Oxford U Press.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14–31. <https://doi.org/10.1016/j.ijhcs.2014.09.006>

Selinger, E., & Whyte, K. (2011). Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture. *Sociology Compass*, 5(10), 923-935. <https://doi.org/10.1111/j.1751-9020.2011.00413.x>

Sethi-Iyengar, S., Huberman, G., & Jiang, W. (2004). How much choice is too much? Contributions to 401 (k) retirement plans. Pension design and structure: New lessons from behavioral finance, 83, 84-87.

SGROI, D. (2015). Happiness and productivity: Understanding the happy-productive worker. *SMF-CAGE Global Perspectives*

SGROI, D. (2015). Happiness and productivity: Understanding the happy-productive worker. *SMF-CAGE Global Perspectives*.

Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review*, 74(3), 433–444.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99. <https://doi.org/10.2307/1884852>

Smith, P. C., Kendall, L. M., & Hulin, C. (1969). The measurement of satisfaction in work and behavior. *Chicago: Raud McNally*.

Spector, P. (1997). *Job Satisfaction : Application, Assessment, Causes, and Consequences*. <https://doi.org/10.4135/9781452231549>

Spector, P. E. (1985). Measurement of human service staff satisfaction : Development of the job satisfaction survey. *American journal of community psychology*, 13(6), 693–713.

Stajkovic, A. D., & Luthans, F. (2001). Differential effects of incentive motivators on work performance. *Academy of management journal*, 44(3), 580-590.

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, 23(3), 165. <https://doi.org/10.2307/2295722>

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 1159-1202.

TechSci Research, R. (2019). *Global Gamification Market By Solution , By Deployment , By Organization Size , By Application , By End-User Vertical , By Region, Competition, Forecast & Opportunities, 2024* (p. 120).

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, 4(3), 199-214.

Thaler, R. (1987). The psychology of choice and the assumptions of economics. *Laboratory experimentation in economics: Six points of view*, 99-130.

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, 112(S1), S164-S187.
<https://doi.org/10.1086/380085>

Thaler, Richard H.; Sunstein, Cass R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press. ISBN 978-0-14-311526-7. OCLC 791403664.

Van Boven, L., Loewenstein, G., Dunning, D., & Nordgren, L. F. (2013). Changing places: A dual judgment model of empathy gaps in emotional perspective taking. *In Advances in experimental social psychology (Vol. 48, pp. 117-171)*. Academic Press.

Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction : How good are single-item measures? *Journal of Applied Psychology*, 82(2), 247-252.
<https://doi.org/10.1037/0021-9010.82.2.247>

Weiss, D. J., Dawis, R. V., & England, G. W. (1967). Manual for the Minnesota satisfaction questionnaire. *Minnesota studies in vocational rehabilitation*.

Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, 61(2), 341-355.
<https://doi.org/10.1111/j.1467-9248.2012.00974.x>

Yellen, J. L. (1984). Efficiency Wage Models of Unemployment. *The American Economic Review*, 74(2), 200-205.

Zúñiga-Vicente, J. Á., Alonso-Borrego, C., Forcadell, F. J., & Galán, J. I. (2014). Assessing the effect of public subsidies on firm R&D investment: a survey. *Journal of Economic Surveys*, 28(1), 36-67.

Three Essays on Using Nudges in Business Firms

In this work, we use the Nudge approach to solve behavioral problems that business firms may have to face. In the first chapter, we start by exposing some issues that the classical economic approach struggles with, before presenting the Nudge approach and why we believe it is relevant to the problems that businesses still face today.

In the second chapter, we change the formulations of the invitations to participate to web surveys, using Nudge principles, in order to improve participation rate. Most Nudges increase the proportion of individuals giving their e-mail address, but only those that acknowledge the respondent's effort increase participation rate.

In the third chapter, we use the Nudge approach's teachings to improve the measurement of job satisfaction. We measure the satisfaction of interns every month during their internships with a very short survey, and compare it to a lengthy survey administered at the end of the internship. We find that satisfaction during the first month of the internship is highly correlated with final satisfaction, which makes it possible to detect potential problems very early.

In the final chapter, we use Nudges to improve productivity by making a simple task more playful, a process called "gamification". Nudges generate the same increase in productivity as the monetary incentives, without the added cost of the latter. Moreover, unless monetary incentives are implemented at the same time, Nudges increase intrinsic motivation.

We conclude our work with practical advices for decision-makers who want to try Nudging.

Keywords: Nudge – Behavioral Economics – Participation rate – Gamification – Job satisfaction – Motivation – Business – Psychology – Marketing

Trois essais sur l'utilisation du Nudge dans les entreprises

Dans cette thèse, nous utilisons l'approche Nudge pour résoudre des problèmes comportementaux auxquelles les entreprises peuvent être confrontées. Dans le premier chapitre, nous commençons par exposer certains des problèmes que l'économie classique peine à résoudre, puis nous présentons l'approche Nudge et expliquons en quoi elle peut être une alternative efficace.

Dans le deuxième chapitre, nous utilisons le Nudge pour changer la formulation d'invitations à participer à des enquêtes sur internet afin d'en augmenter le taux de participation. Nous trouvons que la plupart des Nudges parviennent à convaincre les individus de laisser leur adresse e-mail, mais que seuls ceux qui reconnaissent leurs efforts améliorent le taux de participation.

Dans le troisième chapitre, nous utilisons le Nudge pour développer des enquêtes de satisfaction très courtes, administrées chaque mois, et comparons les résultats ainsi obtenus à ceux d'une enquête plus longue en fin de stage. Nous trouvons que la satisfaction lors du premier mois est fortement corrélée à la satisfaction finale, ce qui rend possible de détecter d'éventuels problèmes très tôt.

Dans le chapitre final, nous utilisons les Nudges pour rendre une tâche simple plus ludique, une approche appelée « gamification ». Les Nudges augmentent la productivité autant que des incitations monétaires, sans coûter autant que ces dernières. Par ailleurs, les Nudges améliorent la motivation intrinsèque quand ils ne sont pas implémentés en même temps que les incitations.

Nous concluons avec des conseils pratiques pour les décideurs souhaitant essayer le Nudge.

Mots-clés : Nudge – Economie comportementale – Taux de participation – Gamification – Satisfaction au travail – Motivation – Entreprise – Psychologie – Marketing