



HAL
open science

Empirical Investigations on the Behavioral Determinants of Political Decisions

Matthieu Pourieux

► **To cite this version:**

Matthieu Pourieux. Empirical Investigations on the Behavioral Determinants of Political Decisions. Economics and Finance. Université de Rennes, 2022. English. NNT : 2022REN1G005 . tel-04236657

HAL Id: tel-04236657

<https://theses.hal.science/tel-04236657v1>

Submitted on 11 Oct 2023

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

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

Thèse de doctorat de

L'UNIVERSITE DE RENNES 1

Ecole Doctorale n°597

Sciences Economiques et sciences De Gestion

Spécialité : Sciences Economiques

Par

Matthieu POURIEUX

**Empirical Investigations on the Behavioral Determinants
of Political Decisions**

Thèse présentée et soutenue à RENNES, le 07 Juillet 2022

Unité de recherche : CNRS UMR-6211 CREM

Rapporteurs avant soutenance :

Mathieu LEFEBVRE Professeur, Aix Marseille School of Economics

Sonia PATY Professeure, Université Lumière Lyon 2

Composition du Jury :

Président : Fabrice LE LEC Professeur, Université de Lille

Examineurs : Mathieu LEFEBVRE Professeur, Aix Marseille School of Economics

Sonia PATY Professeure, Université Lumière Lyon 2

Karine VAN DER STRAETEN Directrice de Recherche, CNRS,
Toulouse School of Economics

Directeur de thèse : Laurent DENANT-BOEMONT Professeur, Université de Rennes 1

Co-directeur de thèse : Matthieu LEPRINCE Professeur, Université de Bretagne Occidentale

The University of Rennes 1 does not intend to endorse nor deprecate any of the views expressed in this Ph.D thesis. These views are to be regarded only as those of their author.

L'Université de Rennes 1 n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

Remerciements

Je tiens tout d'abord à remercier chaleureusement mes deux directeurs de thèse, Laurent Denant-Boémont et Matthieu Leprince, pour leur implication permanente durant ces 5 années de thèse. J'ai eu la chance de bénéficier d'un encadrement à la fois scientifiquement stimulant mais aussi particulièrement humain et bienveillant. Leur disponibilité quasi quotidienne, leur écoute (parfois longue), leur patience, leur soutien (y compris sur des questions personnelles), leurs critiques (presque) toujours constructives, leurs encouragements répétés, leur engagement (y compris pour des virées nocturnes dans l'Ouest profond pour mener à bien des sessions), leur gentillesse et leur professionnalisme m'ont été particulièrement précieux et ont très largement contribué au rendu final.

Je remercie ensuite les membres de mon jury de thèse de me faire l'honneur d'évaluer mes travaux: Mathieu Lefebvre et Sonia Paty, qui ont accepté d'être rapporteurs de cette thèse, ainsi que Fabrice Le Lec et Karine Van der Straeten.

J'adresse aussi des remerciements particuliers à Fabrice Le Lec et Benoît Le Maux pour avoir accepté de suivre mes travaux dans le cadre du Comité de Suivi Individuel (CSI) et pour votre disponibilité même en dehors de ces réunions annuelles. Nos divers entretiens ont toujours été riches d'enseignements pour moi, notamment pour trouver le recul nécessaire sur le quotidien de la thèse et mieux identifier les qualités comme les défauts du travail engagé, et de surcroît très agréables.

Cette thèse a bénéficié du concours financier de trois institutions sans lesquelles je n'aurais tout simplement pas pu être doctorant, et je les en remercie grandement : la région Bretagne, l'institut de la Caisse des Dépôts et des Consignations (CDC) pour la Recherche, et France Urbaine. Je remercie à ce titre toutes celles et ceux qui ont participé aux réunions régulières de suivi de la thèse, et en particulier Is-

abelle Laudier (CDC), Lucie Renou (CDC), Franck Claeys (France Urbaine), Olivier Landel (France Urbaine) et Claire Delpech (ADCF). Ces temps d'échanges étaient toujours plaisants, intéressants, et m'ont permis d'aborder mes travaux de recherche d'un point de vue différent et complémentaire à la vision quotidienne du chercheur. De plus, les sessions expérimentales ont bénéficié de diverses sources de financement pour la rémunération des sujets que je tiens à remercier: les laboratoires CREM et AMURE (Université de Bretagne Occidentale), le projet POPSU Brest, et le PUCA.

Je remercie aussi tous les volontaires qui ont accepté de participer aux sessions expérimentales constituant le matériau principal de cette thèse. En particulier, je remercie tous les participants élus qui ont dégagé du temps pour faire quelque chose qui leur a souvent paru incongru, voire franchement bizarre, et qui se sont ensuite souvent montrés très intéressés et encourageants. Un grand merci aussi aux divers secrétariats, DGS et Présidents des intercommunalités où je me suis rendu dans le cadre des sessions de terrain pour leur intérêt et leur aide, et plus spécifiquement à Madame Gaëlle Andro de Rennes Métropole pour sa disponibilité et sa réactivité. Pour les sessions ayant eu lieu au laboratoire, je remercie la section 35 de l'Association des Maires de France pour son soutien, sa flexibilité et son aide, et tout particulièrement à sa directrice Madame Isabelle Seven-Maillard pour ses efforts répétés malgré certaines difficultés techniques et organisationnelles.

Je remercie également l'Université de Rennes I et le CREM de m'avoir accueilli et de m'avoir permis de travailler dans d'excellentes conditions. J'y ai fait énormément de rencontres stimulantes avec des personnes d'horizons différents qui m'ont toutes apporté quelque chose, du simple conseil à la discussion de fin d'après-midi qui finit en apéro, en passant par la question en séminaire qui re-dessine complètement les contours de la thèse. Je tiens particulièrement à remercier plusieurs personnes qui ont beaucoup compté. Tout d'abord Maïté Stephan, pour trop de choses pour être exhaustif: le coaching régulier, le bureau toujours ouvert pour quand y'en avait marre ou juste pour parler enfants, les conseils quand mon cerveau partait faire un tour, le template de thèse sous LaTeX, les cafés gratuits, et j'en oublie. Une attention toute particulière pour Elven Priour, pour son aide infiniment précieuse avec Z-tree, Orsee, les Regex ou Camstudio, et pour mener à bien les sessions expérimentales, même lorsque c'était le désastre à 300 kilomètres de distance (damn routeur...). Mais aussi pour toutes nos discussions personnelles à des moments plus ou moins

aléatoires, avec ou sans café, avec ou sans Bó Bùn. Olivier L’Haridon, pour ses multiples conseils avisés et sa disponibilité, pour m’avoir autorisé à utiliser ses tablettes sur le terrain et pour m’avoir donné l’opportunité de présenter à l’Ecole TEPP. Fabien Moizeau, pour nos (longues) discussions enrichissantes sur de nombreux sujets, pour les échanges d’idées et de papiers, et pour m’avoir permis de faire des TD dans un cadre idéal pendant 5 ans. Aurélie Bonein, pour son écoute, ses conseils en économétrie, et nos échanges de parents en galère pendant le Covid. Cécile Madoulet, pour sa sollicitude et son calme face à un doctorant souvent un poil trop stressé. Anne-Marie Pouliquen, pour avoir simplifié les choses à de multiples reprises. Mais aussi Hélène Coda-Poirey, Romain Espinosa, Lucie Germain, Chantal Guéguen, Sabrina Hammiche, Julie Ing, Anne L’Azou, Hélène Le Forner, Thomas Le Texier, Youenn Lohéac, Naïla Louise-Rose, Eric Malin, David Masclet, Franck Moraux, Aurélie Passilly, Thierry Pénard, Benoît Tarroux. Je remercie aussi l’école doctorale EDGE et tout particulièrement Hélène Jean pour sa bienveillance, sa patience (notamment vis-à-vis des rendus pour les différentes deadlines) et aussi son soutien psychologique dans les moments compliqués.

Et je remercie tous les autres doctorants! D’abord évidemment Etienne et Martina, et ce projet un peu f(1)ou qui un an après, et quelques centaines d’échanges par mails, vocaux, Zoom, WhatsApp et autres fait un chapitre, un working paper et énormément de nouvelles connaissances. C’était vraiment chouette, et maintenant y’a plus qu’à! Une petite pensée pour Madeg, qui a commencé en même temps qu’Etienne et moi, et qui maintenant vogue vers d’autres horizons gestionnaires (well done!). Bien sûr, je pourrais en rajouter sur Etienne, mais il vaut mieux qu’on s’appelle, comme d’habitude. Et puis il y a Roberto, mon co-auteur italien préféré (le seul aussi). Il faudra qu’un jour on couche sur papier toutes nos idées. Mais d’abord on a un papier à terminer (ou à commencer, je ne sais plus trop, mais on a déjà de super données!). Et Thibaut. Que dire, après je vais encore le traumatiser. Ah si: merci pour le snow, le café, les discussions méthodo, l’anglais, l’aide sous Stata (ah non, c’est vrai), et de ne pas encore m’avoir bâillonné. Et puis aussi tous les autres: Louise, Shervin, Amaury, Sebastian, Lucile, Enora, Nathalie, Romain, Ewen, Jacques.

Merci à ma famille et à ma belle-famille, qui m’ont soutenu de différentes manières pendant ces longues années, même sans trop comprendre ce que je faisais ou com-

ment m'aider. Merci aussi d'avoir subi ma prose à multiples reprises. Merci à mes amis, d'avoir été là pour me prendre comme je venais et me rendre dans un meilleur état, et d'avoir compris que le "et cette thèse alors ?" c'était juste un prétexte pour sortir Les Aventuriers du Rail et autre Citadelle et boire un coup: Fred et Marion (et Théo), Antoine et Laura (et Pauline et Arthur), Simon et Emilie (et Isaac et Samuel), Sylvain et Gabrielle (et Bastien et Noémie), Adrien et Suzanne (et Emile et ?), Marie et Antho (et Mahé et Nils), Fabien et Caro (et Gabriel), Zil et Franck (et Mignon), Boîte et Marion (et Qwertee), PEB, les Prépas (Claire, Marion, Hélène, Boris, Juliette, Réjane).

Merci à Eliott et à Maya, pour m'avoir en permanence rappelé ce qui était réellement important face aux galères de la thèse, aussi importantes puissent-elles paraître. Et pour avoir m'avoir réconforté tous les jours par vos gestes, vos mots et vos exploits parfois un peu maladroits. Et surtout, merci à Lou. Pour tout. Pour m'avoir suivi en Bretagne et être restée jusqu'à la fin (on est bien). Pour le soutien au quotidien, même quand on a l'impression que ça n'avance à rien (en fait si, regarde !). Pour ta confiance en permanence, même quand la mienne se faisait la malle, et pour les rires, les chants, et la danse. Pour avoir su trouver les mots quand ça allait mal, et me supporter quand ça allait bien. Et l'inverse. Pour toutes les solutions, souvent meilleures que les miennes (mais chut!). Bref, merci d'avoir surmonté tout ça avec moi!

Enfin, merci à tous les autres que je n'ai pas cités ici mais qui ont néanmoins joué un rôle : tous ceux qui m'ont écouté patiemment en conférence, en présentiel comme en distanciel, les joueurs de Go du lundi soir et nos échanges sur la thèse pendant les parties, la personne random qui demande par politesse "et toi tu fais quoi dans la vie ?" et qui ne s'attendait pas à ça, les restaurateurs autour de la fac, Stack Overflow/Exchange, celui qui aura lu ces remerciements jusque-là (+1 s'il a ri, +N s'il lit la suite, avec N très grand).

Résumé

Les décisions quotidiennes des élus ont un impact substantiel sur le bien-être des populations, notamment en ce qu'ils déterminent la fourniture et le financement des biens et services publics pour l'ensemble des citoyens. Par conséquent, la compréhension des motivations et des mécanismes présidant à ces décisions est au cœur de nombreux programmes de recherche au sein de différentes disciplines scientifiques.

La théorie économique analyse traditionnellement les décisions publiques comme résultant de la confrontation entre une demande de la part de la population et une offre de la part des décideurs publics (principalement les élus). Dans la grande majorité des cas, les travaux existants reposent sur un cadre micro-économique fondé sur une hypothèse de rationalité d'une part (i.e. les individus prennent des décisions cohérentes entre elles) et une hypothèse d'opportunisme individuel d'autre part (i.e. les individus agissent avant tout pour améliorer leur bénéfice personnel). Ainsi, les citoyens demandent des politiques publiques qui maximisent leur gain personnel, ce à quoi les élus (maximisant eux-aussi leur intérêt personnel) réagissent rationnellement. Cette double hypothèse a néanmoins été largement remise en question par les travaux d'économie comportementale qui intègre à l'analyse économique les apports d'autres disciplines scientifiques (principalement de la psychologie). Ces travaux soulignent en effet l'existence de nombreuses déviations par rapport à la stricte rationalité ainsi que la pertinence de motivations sociales dans la prise de décision (e.g. un concernement pour autrui ou pour la justice des décisions).

L'approche comportementale a connu un gain d'intérêt croissant au cours des vingt dernières années, notamment dans l'élaboration des politiques publiques. Par exemple, prenant acte du fait que les agents économiques sont susceptibles de faire des erreurs, de nombreuses politiques publiques ont intégré des *nudges*, ou incita-

tions non-monétaires, afin d'aider les individus à prendre de meilleures décisions. De nombreux travaux ont aussi permis d'identifier comment des préférences "non-standards" (c'est-à-dire non-égoïstes) influencent la demande de politiques publiques ou le fonctionnement de la démocratie. Cependant, cette approche n'a été que peu mobilisée dans la compréhension même des décisions des élus: elle a davantage servi à alimenter une réflexion sur l'adaptation des politiques publiques à des agents irrationnels ou non-égoïstes qu'elle n'a été appliquée aux agents qui mettent directement en œuvre ces politiques. Autrement dit, les déterminants comportementaux des décisions des élus restent encore méconnus.

Cette thèse mobilise des apports de l'économie politique et de l'économie comportementale et expérimentale dans un cadre micro-économique pour proposer trois études concernant les déterminants comportementaux des décisions publiques. Les deux premiers chapitres se concentrent sur les déterminants des décisions des élus en tant qu'individus. En particulier, nous évaluons le degré de rationalité des décisions prises par des élus en laboratoire dans le chapitre 1 et nous estimons leurs préférences distributives au niveau individuel dans le chapitre 2. Le chapitre 3 adopte une perspective plus agrégée en proposant d'évaluer le lien empirique entre les préférences individuelles des populations (propension à prendre des risques, patience, altruisme et confiance) et les décisions publiques prises pendant la crise sanitaire liée au coronavirus. A ce titre, nous observons au niveau des pays les facteurs comportementaux permettant d'analyser la prise de décision publique.

Sur un plan méthodologique, cette thèse mobilise des données d'expérimentation économique en laboratoire, tout particulièrement dans les chapitres 1 et 2. Ces données présentent plusieurs avantages, notamment par comparaison à des données d'observation ou d'enquête. Premièrement, une expérimentation économique permet d'obtenir des informations de première main sur le comportement individuel et d'inférer le mécanisme sous-jacent aux décisions. En effet, les données d'observation ne permettent souvent pas d'isoler finement le décideur et ses motifs. C'est en particulier le cas pour des décisions impliquant de nombreux acteurs et une règle de décision collective. Deuxièmement, une expérimentation économique permet de confronter les décideurs à des scénarios simplifiés permettant de gagner en précision explicative. Troisièmement, une expérimentation de laboratoire permet de varier l'environnement de décision aisément ainsi que d'obtenir de multiples décisions pour

un même individu pour un coût limité, ce qui demeure complexe avec d'autres types de données. Quatrièmement, la méthodologie expérimentale en économie repose sur des fondements d'anonymat et d'incitations monétaires qui garantissent la crédibilité des réponses obtenues. Ces avantages sont évidemment contrebalancés par la question de la pertinence des mesures à l'extérieur du laboratoire (ou "validité externe"). Mais c'est en partie l'enjeu du chapitre 3 que d'établir un lien entre les décisions publiques effectives et les mesures expérimentales au sein de la population.

Chapitre 1 : Les élus sont-ils efficaces dans leur prise de décision ? Une approche par les préférences révélées

Le chapitre 1 propose un test empirique de la validité du modèle de choix rationnel dans l'explication du comportement des élus. Pour ce faire, nous avons recruté un échantillon d'élus locaux français pour participer à une expérimentation économique. Cet échantillon est constitué de participants à une expérimentation de laboratoire au LABEX-EM (laboratoire d'économie expérimentale de l'Université de Rennes 1) et de participants à une expérimentation de terrain au sein d'EPCI bretons par l'intermédiaire d'un laboratoire mobile. Nous comparons le comportement des élus avec celui observé au sein de deux groupes de contrôle: un groupe de contrôle étudiant et un groupe de contrôle issu de la population française et dont les caractéristiques socio-démographiques sont comparables à celles des élus. De telles comparaisons permettent d'identifier dans quelle mesure les élus adoptent un comportement singulier.

Chaque participant à l'expérimentation est confronté à une série de jeux du dictateur où il doit allouer une somme d'argent entre lui-même et un autre participant (Andreoni and Miller, 2002, Choi et al., 2007). A l'aide de techniques de préférences révélées, nous évaluons ensuite la proximité de chaque participant avec le modèle de choix rationnel, c'est-à-dire sa propension à prendre des décisions qui sont conformes à un ordonnancement cohérent des préférences. Plus spécifiquement, nous évaluons la proximité des données de chaque participant avec l'axiome généralisé des préférences révélées (GARP) dont il a été démontré l'équivalence avec la rationalisation des données par une fonction d'utilité bien définie (Afriat, 1967, Varian, 1982). Les participants dont les décisions sont conformes avec l'axiome sont considérés "rationnels" ou "efficaces", dans le sens où il ne gaspillent pas de ressources

en prenant des décisions contradictoires entre elles. Afin d'intégrer la possibilité de dévier de l'axiome de différentes façons, nous calculons trois indices de rationalité fréquemment mobilisés dans la littérature: l'indice d'efficacité d'Afriat (Afriat, 1972), l'indice d'efficacité de Houtman et Maks's (Houtman and Maks, 1985) et le vecteur d'efficacité de Varian (Varian, 1990, 1993). Les deux premiers indices sont calculés sur la base d'algorithmes simples tandis que le dernier est calculé sur la base de la résolution d'un programme linéaire mixte. Globalement, ces indices permettent de distinguer la taille des déviations du nombre de déviations, ainsi que de distinguer la déviation agrégée des déviations causées par chaque décision. A cet égard, notre analyse se différencie des études existantes qui se focalisent généralement sur un seul indice de rationalité (le plus souvent celui d'Afriat).

Nos résultats indiquent que les élus participants s'écartent du modèle de choix rationnel, mais pas de façon substantielle. En particulier, et bien qu'ils soient plus éloignés de la rationalité que les étudiants, leurs indices de rationalité ne sont pas différents de ceux calculés pour une population comparable. Nous retrouvons des résultats similaires dans le cadre d'une analyse de régression et nous effectuons de plus des analyses de puissance *ex-post* qui confortent nos interprétations. Ces résultats témoignent d'un côté de la pertinence empirique du modèle du choix rationnel pour expliquer les décisions prises par des élus (locaux). D'un autre côté, ils mettent aussi en exergue des similitudes avec d'autres populations dans les mécanismes de la prise de décision. Ceci implique que les biais cognitifs observés au sein d'autres populations sont aussi susceptibles d'être observés chez les élus. Cette perspective pourrait alors conduire à d'autres travaux sur la sensibilité des élus à ces biais et à une réflexion plus large sur les possibilités de les éviter.

Chapitre 2 : Les préférences distributives des élus : mesures expérimentales

Le chapitre 2 évalue l'importance des motivations intrinsèques pour les élus, et par extension la pertinence empirique de l'hypothèse de maximisation de l'intérêt personnel pour cette population. Plus spécifiquement, notre analyse porte sur les préférences des élus en matière de répartition non stratégique de ressources économiques (ou "préférences distributives") et distingue deux arbitrages: l'arbitrage entre Egoïsme et Altruisme, et l'arbitrage entre Egalité et Efficacité. Le premier ar-

bitrage oppose la maximisation d'un gain personnel avec la maximisation d'un gain pour autrui. Le second arbitrage oppose la minimisation des inégalités et la maximisation de la somme des gains.

Notre stratégie empirique repose sur une procédure en quatre étapes qui mobilise les mêmes données expérimentales que celles utilisées dans le chapitre 1 et qui est essentiellement dérivée de la littérature existante (et tout particulièrement: Andreoni and Miller, 2002, Fisman et al., 2007). Premièrement, nous déterminons les demandes individuelles pour la distribution de ressources par la résolution d'un programme de maximisation du consommateur fondé sur une utilité de type CES qui capte les deux arbitrages. Deuxièmement, nous trions les données individuelles sur la base de leur proximité avec un ordonnancement homothétique des préférences évaluée avec différents indices d'homothétie (Heufer and Hjertstrand, 2015, 2019). Cette seconde étape est cruciale pour différencier les préférences réelles d'un comportement irrationnel. En comparaisons de la littérature, qui en majorité effectue un tri des données sur la base de la seule rationalité, nous tenons compte des déviations par rapport à l'homothétie, sur laquelle la forme fonctionnelle CES est implicitement fondée. En définitive, nous exposons les participants à un test des préférences révélées qui est plus exigeant que les tests utilisés usuellement dans la littérature. Troisièmement, nous estimons structurellement les demandes individuelles pour les données suffisamment proches de l'homothétie afin d'obtenir deux paramètres correspondant à chacun des deux arbitrages. Quatrièmement, nous comparons les distributions de ces paramètres entre nos groupes expérimentaux à l'aide de tests statistiques usuels et d'analyses de régression.

Nos résultats indiquent que les élus accordent davantage d'importance à autrui et à la réduction des inégalités que nos deux groupes de contrôle, qui sont plus égoïstes et sensibles à l'efficacité. Ces résultats sont robustes à une batterie de tests de robustesse, et en particulier à la différence entre les groupes de proximité avec l'homothétie, à la différence entre les sessions de laboratoire et les sessions de terrain au sein des élus, et à l'inclusion de différents contrôles socio-démographiques. Ces résultats soulignent l'importance des motivations intrinsèques pour les élus relativement à d'autres groupes, et en particulièrement un goût pour l'égalisation, ce qui a des conséquences à la fois théoriques, pour une modélisation plus adaptée des décisions publiques, mais aussi pratiques, pour l'élaboration d'institutions encadrant

efficacement le comportement des décideurs.

Chapitre 3 : L'importance des préférences ! Le lien entre les politiques publiques face à l'épidémie de COVID-19 et les préférences de la population

Le chapitre 3 propose une évaluation empirique de la relation entre des décisions politiques effectives et des préférences non-standards au sein de la population. Nous testons l'hypothèse selon laquelle certaines préférences économiques de la population - propension à prendre des risques, patience, altruisme et confiance - sont corrélées avec la réactivité du gouvernement et l'intensité des politiques publiques durant la crise sanitaire du COVID-19 au sein d'un échantillon de 109 pays. Notre analyse empirique combine des données mondiales sur les politiques publiques nationales fournies par l'Oxford COVID-19 Government Response Tracker (Hale et al., 2020), et des données sur les préférences comportementales au niveau individuel en provenance soit du Global Preferences Survey (Falk et al, 2018) soit des 6^{èmes} et 7^{èmes} vagues du World Values Survey (Inglehart et al. 2014, Haerpfer et al, 2020). Nous menons des analyses de régression en coupe transversale (sur la réactivité) et en panel (sur l'intensité) en intégrant les préférences comportementales comme variables explicatives pour identifier des corrélations.

Nos résultats indiquent que la confiance apparaît comme un déterminant important des deux aspects de la réponse politique. Premièrement, les gouvernements de pays où le niveau de confiance est élevé ont attendu plus longtemps avant de mettre en œuvre leur première politique restrictive. A l'inverse, les autres préférences comportementales (propension à prendre des risques, patience et altruisme) ne semblent pas liés à la réactivité des gouvernements. Deuxièmement, toutes les mesures de préférences sont liées positivement à l'intensité des politiques publiques à moyen et/ou long terme, bien que ces liens soient sensibles à l'horizon temporel considéré. A court terme, seules la patience et la confiance sont (négativement) liées à l'intensité des politiques publiques. Ces résultats sont robustes à des modifications de la définition de la réactivité, à l'inclusion de variables de contrôle au niveau national, à l'utilisation de différentes bases de données, et à une analyse séparant les continents. Bien que ne permettant aucune interprétation causale stricte, ces résultats soulignent l'importance de certaines préférences à un niveau agrégé dans un con-

texte de crise, ce qui peut avoir des conséquences notamment sur la modélisation du comportement à l'origine des décisions publiques.

Contents

General Introduction	1
1 - Are Politicians Efficient Decision-Makers? A Revealed Preference Approach	17
1 Introduction	18
2 Related literature and behavioral conjectures	25
3 Experimental design	33
3.1 Subject pools and procedures	33
3.2 The experimental task	42
4 Revealed Preference Analysis	44
4.1 The Generalized Axiom of Revealed Preferences (GARP)	44
4.2 Rationality measures	45
5 Results	48
5.1 Between-groups comparisons on rationality	48
5.2 Regression framework	53
6 Power analyses	57
6.1 Power of the revealed preference tests	57
6.2 Power of the statistical tests	61
7 Concluding remarks	62
References	66
Appendix A - Are Politicians Efficient Decision-Makers?	79
A.1 Complements on experimental procedures	80
A.1.1 Instructions	80
A.1.2 Brief description of the subsequent experimental tasks	86
A.1.3 Questionnaire and definition of the socio-demographic variables	89
A.1.4 Sessions' information	94
A.2 Complementary revealed preference analysis	97

A.2.1	Computation of the indices	97
A.2.1.1	The Afriat (1972, 1973)'s Efficiency Index (AEI) . . .	97
A.2.1.2	The Houtman and Maks (1985)'s Index (HMI) . . .	98
A.2.1.3	The Varian (1990, 1993)'s Efficiency Vector (VEV) .	100
A.2.2	Additional results on between-groups comparisons	103
A.2.3	Inconsistencies over time	105
A.2.4	Caveat with the use of AEI/VEV	109
A.3	Complementary regression analyses	114
A.3.1	Effect of past experience in experiments	114
A.3.2	Heterogeneity analysis of rationality within Politicians	118
A.4	Complementary power analyses	121
A.4.1	Power of the revealed preference conditions	121
A.4.2	Power of the statistical tests	125
References	128

2 - The Distributional Preferences of Politicians: Experimental

Evidence		131
1	Introduction	132
2	Literature-based behavioral conjectures	140
2.1	Selfishness/Altruism tradeoff	141
2.2	Equality/Efficiency tradeoff	146
2.3	Conjectures	149
3	Experimental design	153
3.1	Subject pools and procedures	153
3.2	The modified dictator game	155
4	Distributional preferences analysis	159
4.1	Revealed preference analysis: consistency with homothetic preferences	159
4.2	Theoretical framework and empirical strategy	161
5	Results	166
5.1	Raw data overview	166
5.2	Between-groups comparisons in estimated parameters	171
5.2.1	Selfishness/Altruism tradeoff	172
5.2.2	Equality/Efficiency tradeoff	176

5.2.3	Additional analyses and robustness checks on estimated parameters	180
5.3	Regression framework	182
6	Discussion	189
7	Concluding remarks	194
	References	198
Appendix B - The Distributional Preferences of Politicians		213
B.1	Complementary revealed preference analysis	214
B.1.1	Definitions of homothetic indices	215
B.1.1.1	The Homothetic Efficiency Index (HEI, Heufer, 2013)	215
B.1.1.2	The Homothetic Houtman & Maks Index (H ² MI, Heufer and Hjertstrand, 2015)	217
B.1.1.3	The Homothetic Efficiency Vector (HEV, Heufer and Hjertstrand, 2019)	217
B.1.2	Between-groups comparisons in homothetic indices	219
B.1.3	Closeness to homothetic preferences: power computations	224
B.2	Complements on the empirical strategy	226
B.2.1	Description of the structural estimation process	226
B.2.2	Description of the estimation restrictions	229
B.2.2.1	Subjects-based restrictions	229
B.2.2.2	Observations-based restrictions	233
B.3	Complementary analyses on group comparisons from raw data	235
B.3.1	Robustness check on cross-sessions contagion effect	235
B.3.2	Power of the statistical tests comparing Field and Lab on raw data	239
B.4	Complementary analyses on group comparisons from parameters	243
B.4.1	Distributions of estimated parameters for each restriction strategy	243
B.4.2	Correlations between estimated parameters	250
B.4.3	Comparisons with Fisman et al. (2007)'s data	252
B.4.4	Distributions of prediction errors	254
B.5	Complementary regression analyses	263
B.5.1	Specification curves for the regressions on the estimated parameters	263

B.5.2	Regressions without political preferences in Field v. Non-Students	269
B.5.3	Secondary results on control variables	270
B.5.4	Effect of past experience in experiments	275
B.5.5	Alternative measures of political preferences	277
B.5.6	Heterogeneity analysis of estimated parameters within Politicians	283
	References	287
3 - Preferences matter! Political Responses to the COVID-19 and Population's Preferences		293
1	Introduction	294
2	Data and measures	296
2.1	Political responses	297
2.2	Preferences	298
3	Results	300
3.1	Governments' <i>Responsiveness</i>	302
3.2	<i>Stringency</i> of the political responses	305
4	Concluding remarks	310
	References	312
Appendix C - Preferences matter!		317
C.1	Data Sources	318
C.1.1	The Oxford COVID-19 Government Response Tracker (OxCGRT)	318
C.1.2	The Global Preferences Survey (GPS)	321
C.1.2.1	Content of the survey	321
C.1.2.2	Item-selection procedure	323
C.1.3	The World Values Survey (WVS)	326
C.1.4	Country-level control variables	328
C.1.5	List of countries	329
C.2	Descriptive statistics	334
C.2.1	Statistics on dependant variables	334
C.2.2	Relationships between independant variables	339
C.3	Model specification	342

CONTENTS

C.3.1	OLS regression on <i>Responsiveness</i>	342
C.3.2	Fixed effect panel regression on <i>Stringency</i>	343
C.4	Robustness checks	344
C.4.1	<i>Responsiveness</i> : alternative datasets and definitions	344
C.4.2	<i>Responsiveness</i> : Full regression tables	347
C.4.3	<i>Stringency</i> : Additional results using WVS data	351
C.4.4	Analysis by continents	354
	References	359
	General Conclusion	361
	Stata References	367

General Introduction

"A large part of the research effort is necessarily devoted to observation and experiment ; simply to finding out how things are in fact"

Herbert Simon, Rationality in political behavior,
Political Psychology, 1995

The decisions made by political actors on a daily basis affect citizens' welfare in a substantial way. In democratic countries it is even the very purpose of political representation to enhance the nation's prosperity. The share of public spending in the GDP, which may serve as an indicator of the relative importance of the public sector as compared to the private sector, is considerable in many countries. Governments' global expenses represented 27.9% of the world GDP in 2019, a share which is even more important in developed countries (e.g., 36.6% in the European Union and 46.3% in France).¹ During the course of their every day life, citizens regularly experience the results of political decisions (schools, hospitals, police, justice, public transportation, garbage collection, parks and so on). Governments' intervention is also visible directly on the payroll due to taxes and redistribution systems. As a consequence, understanding the drivers of such political decisions has been perceived as a fundamental research question for many decades and in various scientific disciplines.

¹The data was obtained from the World Bank (<https://data.worldbank.org/indicator/GC.XPN.TOTL.GD.ZS?view=chart>), building on multiple sources (International Monetary Fund, Government Finance Statistics Yearbook and data files, and World Bank and OECD GDP estimates). Expense is defined as "cash payments for operating activities of the government in providing goods and services[, which includes]compensation of employees (such as wages and salaries), interest and subsidies, grants, social benefits, and other expenses such as rent and dividends".

Very succinctly, the economic analysis of political decisions have historically been nurtured by two competing though complementary strands of literature, each of which adopts different premises regarding the public decision-making framework. On the one hand, the welfarist approach proposes a normative framework under which a benevolent planner is responsible for political decisions. Such planner is endowed with a social welfare function which he attempts to maximize. The fundamental question is therefore to determine the shape and inputs of the social welfare function (see Mueller, 2003, for a review of the diversity of approaches in this area). On the other hand, the public choice approach proposes a positive framework under which political decisions are made by rational and self-oriented decision-makers (see again Mueller, 2003, for an extensive review). Specifically, such approach addresses a fundamental critique to the welfarist approach: it argues that the governments are not omniscient benevolent planners but are rather composed of people who are endowed with their specific motivations. Assuming that people in general behave out of their personal interest, we must subsequently treat policy-makers as self-interested utility maximizers and model their behavior accordingly. This statement is developed out of a "behavioral symmetry" argument (Buchanan and Tullock, 1962), according to which there is no valid reason for research studies to treat politicians differently from the remainder of the population because, in the end, they are people too. Ignoring such a feature contributes to a biased analysis of policy-making which cannot explain a wide range of stylized facts (e.g., rent-seeking behaviors, pork-barrel politics, differing size and organizational structure of the government, or even the mere congruence with the median voter).

However, the standard microeconomic approach, which consists of the combination of a cognitive (rationality) and a motivational (self-interest) assumption about politicians' behavior, appears quite restrictive. From today's scientific perspective, maintaining these two assumptions as natural hypotheses for an economic analysis of political decisions even appears deprecated. Indeed, much of the work in behavioral economics, supported by countless experimental findings, has been devoted to underlying that people regularly fail to maximize their utility or deviate from strictly egoistic goals to incorporate others in their decision-making process (DellaVigna, 2009).

Interestingly, the notion that people are likely subject to behavioral biases has been recently advanced as a novel rationale for public intervention, alongside stan-

dard market failures (low-competition environments, externalities, asymmetries of information and public goods). The basic idea is that, in order to achieve an optimal allocation of resources, the government should help people overcome their cognitive limitations. Such a reasoning led to the vast literature on "nudges" and the emergence of nudge units which influence governments throughout the world.² Yet, maintaining in the same framework that the people who implement such public intervention behave rationally constitutes a paradox (Viscusi and Gayer, 2015). If one wants to take seriously the "behavioral symmetry" critique according to which all people should be treated equally, then one should embrace it and question the rationality assumption for politicians as well.

A similar argument could be made for non-selfish preferences. Since such preferences are fundamental drivers of behavior in many different situations, designing a policy to tackle a specific issue requires to take into account such preferences. Otherwise the policy may miss the target and ultimately waste economic resources even though the objective was to improve welfare.³ But arguing that non-standard preferences must be taken into account while designing policies on the one hand and maintaining that policy-makers are necessarily selfish when analyzing their behavior on the other hand again does not resist the "behavioral symmetry" critique.

Furthermore, to date and to the best of our knowledge, both assumptions about politicians' behavior (rationality and selfishness) have not been submitted to a stringent empirical test based on individual observations. More generally, the incorporation of behavioral insights into the analysis of political decisions is gaining interest over the recent years but remains particularly scant (see e.g., Wilson, 2011, Schnellenbach and Schubert, 2015, Harstad and Selten, 2016, Zamir and Sulitzeanu-Kenan,

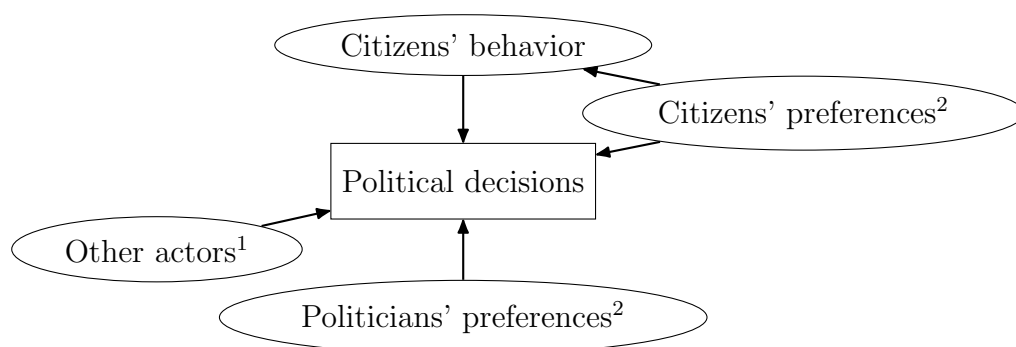
²Nudges are generally defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler and Sunstein, 2008). Typical nudges include modifying order of dishes, green-to-red light bulb on energy consumption, changing baseline option for post-mortem organ donation, target flies in urinoirs, etc. One example of nudge unit in France is the BVA Nudge Unit, whose expertise has been called for repeatedly by the French government during the COVID-19 epidemic.

³For a recent example, Alfaro et al. (2020) incorporate social preferences in a SIR (Susceptible Infected Recovered) model of epidemics to demonstrate that the necessity and effectiveness of emergency sanitary responses crucially relies on such preferences. In particular, since patient and altruistic individuals spontaneously adopt disease-mitigation behavior like reducing mobility, the population's endogeneous reaction to an epidemic may be sufficient to partly address the issue. Consequently, designing a policy which assumes selfish preferences only is likely to waste resources on unnecessary non-pharmaceutical interventions, hence decreasing welfare.

2018). This dissertation intends to contribute to this burgeoning literature by adopting a fundamentally positive and empirical stance.

The general framework for understanding political decisions which is adopted in the present dissertation can be sketched as in the figure I.1.⁴ Policies are conceived as the product of a market interaction between a demand from citizens and a supply from politicians (Persson and Tabellini, 2002).

Figure I.1: Schematic representation of political decisions' drivers



¹ Other actors represent all actors that may have an influence on the political decision-making process (e.g., lobbies, administrative staff, experts, etc.)

² Such preferences encompass a wide array of policy-relevant preferences, including ideological preferences, selfish motives, other-regarding preferences (including distributional preferences), etc.

On the demand-side, citizens express their preferences through various voice mechanisms (in particular by casting a vote, but also more largely through polls, contests, etc.). In democracies, due to the electoral system, politicians are concerned about getting (re)elected and consequently elaborate policy platforms which

⁴Obviously, such a schematic representation is overly simplistic. In the reality, the whole process is endogenous, so that there should also be reverse arrows representing for instance the influence of political decisions on citizens' behavior or preferences. In addition, all relationships which likely exist are not depicted. For instance, the other actors on the supply side may also shape the preferences of either citizens or politicians, through for instance intense lobbying efforts, novel scientific evidence or by contributing to the selection of politicians with specific desirable traits.

converge towards the median voter's preferences (as in the median voter model, see Hotelling, 1929, Downs, 1957) or at least towards what they believe the median voter wants (as in the probabilistic median voter model, see Lindbeck and Weibull, 1987, Coughlin, 1992). Citizens' preferences may also have an indirect effect on political decisions through behavior, which conditions both the necessity and the effectiveness of policies. For instance, a naturally risk-averse population may require a less stringent set of rules and incentives to avoid risky behaviors (overspeed, drug use, junk food, risky investments, etc.) than a naturally risk-loving population.⁵

On the supply-side, politicians also define policy platforms based on their own preferences. In practice, politicians indeed hold an important amount of discretionary power in the political decision-making process, which allows them to prioritize their own goals.⁶ What constitutes such goals is however unclear. The literature has so far contributed to identifying several possible objectives, and principally: policy preferences or "ideologies" (Wittman, 1977, Besley and Coate, 1997), rent-seeking opportunities (Migué et al., 1974, Breton and Wintrobe, 1975), and bureaucratic incentives (Niskanen, 1971, Brennan et al., 1980). But in any case, the primary motive behind all these motives is the maximization of politicians' private benefits. The present dissertation focuses first on the idea that politicians are ultimately good at meeting some goals (i.e., "rational" decision-makers), and second scrutinizes another set of possible goals: non-selfish motives (in particular distributional preferences).

The figure I.1 also emphasizes that investigating the behavioral motives behind actual policy decisions made by politicians is a particularly complex enterprise. Indeed, such decisions involve a very large set of actors (politicians but also citizens, lobbies, administrators, experts, and so on) whose behaviors are aggregated in a way which is difficult to accurately apprehend. Disentangling what an individual politician is ultimately responsible for from such data is therefore not straightforward. Furthermore, it is difficult to isolate behavioral motives from one another

⁵Such a line of thought will be briefly expanded in the context of the COVID-19 epidemic in the introduction of chapter 3.

⁶We acknowledge that the framework described in the figure I.1 does not readily transpose to other political regimes, and in particular dictatorships. However, the influence of politicians' preferences on political decisions in such regimes is arguably even more important since the political institutions and citizens' control are weaker than in democratic regimes.

with observational data. For instance, the congruence of politicians with lobbyists' interests, which is spontaneously perceived as collusion, may ultimately not relate to politicians' self-interest. Lobbies may indeed affect politicians' perception of what the median voter wants so that political decisions are ultimately aligned with the lobbies' agenda without the necessity for politicians to derive any private gains from such decisions. Besides the decision-making environment may vary from one politician to another politician, or from one decision to another decision, which further complicates the analysis. For instance, a landlord politician is likely to hold a personal interest in decisions affecting the housing market, but may not care much about decisions affecting schooling (say, if he also is single). Accounting for the decision environment would therefore be crucial.

One possibility would be to rely on theoretical modelling to derive empirical predictions to be tested, but such a practice entails to impose much structure on the decision-making process and may not resolve interpretation issues. In addition, naturally-occurring policy choices may not provide researchers with the right kind of data to investigate the relevant questions. For instance, an inquiry regarding the rationality of public decision-making would ultimately require a frequent-enough possibility to make irrational decisions, which may not happen in practice. Another possibility, which is common-practice outside of economics, would be to substitute actual data with survey data or mobilize qualitative methods (e.g., interviews). However, such an empirical strategy appears ill-suited for our research purpose. Indeed, asking questions about past behavior or on hypothetical scenarii means focusing on declarative stakes-free statements. Yet, such stakes are fundamental both to our research questions (in particular the focus on self-interest) and more generally to draw an accurate representation of politicians' daily decisions, which essentially boils down to (re)distributing money. Furthermore, such methodologies may not enable to tackle research questions related to politicians' behavior, because such actors may place an extra value on reputation and distort their answers accordingly.

In order to account for these shortcomings, this dissertation proposes to rely on the experimental economics methodology. Running a laboratory experiment with actual politicians provides the opportunity to design an empirical test which is adapted to the researcher's needs. Indeed, it allows to finely control the decision-making environment and to discard most of the noise which characterizes observational data (i.e., high internal validity). In particular, the environment can be reduced to a

simple and straightforward version of what one wants to observe. At the same time, it provides experimental subjects with a clear incentive scheme which enhances the revelation of genuine behavioral motives through observed decisions. Such a feature enables to gauge the importance of intrinsic motivations/non-standard preferences in a context where they are costly to the decision-maker.

To the best of our knowledge, there only exists a handful of incentivized experiments involving actual politicians, many of which being listed in this dissertation. We document possible explanations for such a fact in both chapters 1 and 2, with a special focus on the scarcity of economic studies. In most cases, existing studies come from other disciplines (political science, public administration) and consequently do not rely on the same methodological tools as the ones which are exposed in this dissertation. As a consequence, our findings complement such body of work, which is importantly multi-disciplinary.

At the same time, the experimental economics methodology also comes at the price of loosing much of the context under which political decisions are made. In other words, the external validity remains an open question. For instance, observing decisions made by politicians in the laboratory does not even guarantee that the same politicians would act in accordance with what they revealed during the experiment in a real-world context. Such an argument may be disputed on both logical⁷ and empirical grounds. There indeed exists a vivid dispute in experimental economics with respect to the consistency of subjects' behavior across various contexts (e.g., a laboratory context and a "real" context, see for instance Harrison and List, 2004, Levitt and List, 2007). Globally, the bulk of studies find encouraging results according to which what is observed at some point in the laboratory is consistent with alternative decision environments or repeated observations. However, there also exists some contradictory findings (e.g., Galizzi and Navarro-Martinez, 2019), so that the dispute remains topical.

The approach which we undertake in this dissertation is somewhat different: we scrutinize the relationship between political decisions and well-known behavioral motives using an unique opportunity provided by the COVID-19 sanitary crisis. Indeed, the epidemic is essentially an exogenous shock which affected all countries in the world. Emergency health policies were decided by most governments in a very

⁷One logical argument could be that decisions in the laboratory also are "real-world" decisions which provide some information as to how people behave (Smith, 1982).

short amount of time, and the scientific community took on the daunting task of collecting as much data as possible on the evolution of the crisis (including policy responses). We use this shock as an opportunity to evaluate the relationship between epidemic-related political decisions and non-standard preferences (risk-taking, patience, trust and altruism) measured with worldwide survey studies a few years before the virus arose. Consequently, and although such an analysis does not allow to disentangle possible mechanisms with precision, it provides some stylized facts on the relevance of a behavioral approach to policy-making. As such, this approach should not be interpreted as a direct answer to the external validity critique for chapters 1 and 2.⁸ On the contrary, the idea here is to determine whether we observe some relationships between actual policy decisions and standard behavioral measures using world data. At the same time, the connection between these relationships and an underlying theoretical framework is much less straightforward than what an experiment affords.

Outline

This PhD dissertation lies at the crossroads between political economy and behavioral/experimental economics, with an important microeconomic basis. It tackles three research questions related to the behavioral determinants of political decisions, each of which being addressed in one of the three chapters. We primarily focus on politicians' preferences from the figure I.1. Specifically, we investigate the behavioral motives behind decisions made by politicians by questioning whether actual politicians abide by a rational choice framework (chapter 1) and by eliciting their intrinsic motivations in a distributive context at the individual level (chapter 2). The chapter 3 adopts a more aggregate perspective to evaluate the empirical relationship between actual political decisions and common non-standard preferences

⁸Such an inquiry would have required for instance to link politicians' behavior in the lab with behavior from the same politicians out of the lab (say votes during municipal councils), which would have raised both methodological and ethical questions.

measured within the population (risk preferences, time preferences as well as social preferences in the form of trust and altruism).⁹

The three chapters are deeply empirical and all rely on experimental data. The data for the chapters 1 and 2 were obtained from experiments with actual politicians designed and conducted expressly for the purpose of this dissertation. Although they do not propose a novel theoretical framework, both chapters are tightly connected to consumer theory. The chapter 1 builds on non-parametric tests of revealed preference axioms, whereas the chapter 2 relies on individual-level estimation of a structural utility function. The data for the chapter 3 combine observational data on national political responses during the COVID-19 epidemic with experimentally-validated survey data from existing pre-epidemic worldwide studies. Contrary to the previous chapters, the chapter 3 focuses on aggregated data on political decisions *in vivo* rather than on individual decisions made by political actors *in vitro*. Furthermore, it is a correlational study which resorts to standard cross-section and panel estimation techniques and which does not provide an explicit link with theoretical underpinnings.

Chapter 1 offers a test of the empirical validity of the rational choice framework in explaining politicians' behavior. We recruit a sample of French local politicians to participate in an incentivized experiment either in the laboratory or in the field (within intermunicipal structures). We compare the behavior of politicians with the one observed in two control groups: a student group and a group selected from the general French population. Such comparisons allow to investigate how politicians differ both from the standard laboratory population and from a socio-demographically-comparable group.

The experiment exposes each subject to a series of decisions where a budget is to be allocated between two accounts, one for himself and one for another randomly-selected subject (Andreoni and Miller, 2002, Choi et al., 2007). We then employ revealed preference techniques to assess each subjects' proximity with the rational

⁹All chapters are extended versions of working papers which are either currently available or soon to be so. The working paper corresponding to the chapter 1 is a single-author work. The working papers corresponding to the chapters 2 and 3 are co-signed with two co-authors: Laurent Denant-Boèmont & Matthieu Leprince (both PhD advisors) for the former, and Etienne Dagorn (PhD) & Martina Dattilo (PhD student) for the latter.

choice framework, that is whether subjects make decisions that are consistent with a well-defined preference ordering (i.e., "rational" or "efficient" decision-makers). Specifically, we assess each individual dataset's closeness to the Generalized Axiom of Revealed Preference (GARP), which is equivalent to the rationalization by some well-behaved utility function (Afriat, 1967, Varian, 1982), by computing three rationality indices. These indices differ in their definition of "closeness" to GARP and thus allow to refine our interpretation of the rational framework's validity. In this regard, our analysis stands out from most of the literature which focuses on one rationality index only.

We find that politicians do differ from the rational choice framework, but not substantially. On average they waste about 11% of their budget by making inconsistent choices and only 2 to 3 decisions are generally responsible for the observed deviations. Overall, the average behavior over all decisions is 99% as efficient as optimizing behavior. Furthermore, although they are farther away from rationality than students (by 5 percentage points), politicians' rationality indices are not different from a comparable control group. In addition, we do not find any statistical difference between our lab and our field sessions. We recover the same results in a regression framework and also run *ex-post* power analyses which consolidate our interpretations. In particular, all our participants are much closer to the rational choice framework than simulated uniformly-random players. Finally, we discuss the theoretical and methodological consequences of these findings in the chapter's conclusion.

Chapter 2 gauges the importance of intrinsic motivations among politicians, thereby assessing the empirical relevance of the selfishness assumption. Specifically, we focus on politicians' preferences over the non-strategic distribution of monetary resources (or "distributional preferences") and disentangle two tradeoffs: the Selfishness/Altruism and the Equality/Efficiency tradeoffs. The former pits maximizing

one's own gains against maximizing the gains of others. The latter pits minimizing inequalities against maximizing the sum of gains.¹⁰

Our empirical strategy consists in a four-steps procedure which makes use of the same experimental data as in the chapter 1 and which essentially borrows from the existing literature (foremost: Andreoni and Miller, 2002, Fisman et al., 2007). First, we theoretically derive the individual demands for distribution as the result of a CES utility maximization program which captures both tradeoffs. Second, we screen individual data based on their proximity to homothetic utility maximization based on various homothecity indices Heufer and Hjertstrand (2015, 2019). This second step is crucial to separate true preferences from irrational behavior. As compared to the literature, which predominantly screens only for GARP-inconsistent choices, we take into account deviations from homothetic preferences, which are entailed by the CES functional form. Ultimately, we expose subjects to a more stringent test of revealed preference conditions than what is standard practice. Third, we structurally estimate the demands at the individual level for homothetically-close-enough data and obtain two parameters which correspond to the two tradeoffs. Fourth, we compare the distributions of the two parameters between our experimental groups using both standard statistical tests and regression analyses.

We find that politicians are much more concerned about others and about reducing inequalities than both our control groups, which are more selfish and more efficiency-oriented. Specifically, politicians favor an uniform distribution: more than 55% of our sample roughly equalizes the payoffs, which is 25 (resp. 35) percentage points more than a comparable (resp. student) group. These results are explained first by a lower concentration of near-selfish subjects and higher concentration of fair-minded subjects (equal weights placed on self and other) among politicians. Second,

¹⁰This dissertation makes use of two different notions of "Efficiency", which may be confusing and thus deserves some preliminary clarifications. The chapter 1 focuses on characterizing the decision-making process of individuals with respect to the utility maximization framework. In this context, an efficient individual is a decision-maker who manages to avoid making decisions which waste resources, thus an individual who is consistent with the utility maximization framework. The chapter 2 focuses on characterizing the underlying preferences of decision-makers. In particular, such preferences may be oriented towards two different conceptions of distributive justice: "Efficiency", understood as the maximization of the sum of monetary gains, or "Equality", understood as the minimization of inequalities. Since these two conceptions each entail a consistent decision pattern, they are both compatible with an efficient decision-making process (or "rational"). In other words, the former definition of "Efficiency" differentiates Pareto-optimal allocations from non-Pareto-optimal allocations, whereas the latter definition characterizes a subset of Pareto-optimal allocations as opposed to another subset composed of "egalitarian" allocations.

politicians are predominantly equality-oriented (two-thirds of our sample), whereas the comparable group is globally balanced with respect to the Equality/Efficiency dilemma and students favor efficiency. These results resist a variety of robustness checks, including most notably differences in closeness to homothetic preferences, differences among politicians between field and lab sessions, and the inclusion of several socio-demographic controls. Such findings thus point towards the relative importance of intrinsic motivations for politicians as compared to other groups, and especially a taste for equalization. Finally, we address the main shortcomings of our experimental design with respect to our results in a discussion section and draw some theoretical as well as practical implications for these results in the chapter's conclusion.

Chapter 3 proposes an empirical assessment of the relationship between actual political decisions and non-standard preferences within the population. We test the hypothesis that populations' economic preferences - risk, time trust and altruism - matter both in terms of the governments' responsiveness and its stringency during the COVID-19 sanitary crisis in a total of 109 countries. Our empirical analysis combines data on worldwide political responses from the Oxford COVID-19 Government Response Tracker (Hale et al., 2020), individual economic preferences from the Global Preferences Survey (Falk et al., 2018), and the 6th and 7th waves of the World Values Surveys (Inglehart et al., 2014, Haerpfer et al., 2020). We run cross-section (on responsiveness) as well as panel (on stringency) regressions with behavioral motives as explanatory variables to identify correlational patterns.

We find that trust is an important driver of both aspects of political responses. First, countries with high levels of trust wait longer before implementing their first policy to tackle the epidemic. On the contrary, other preferences (risk-taking, patience and altruism) do not relate to responsiveness. Second, all measures of preferences positively relate to the intensity of the political responses in the middle-run or in the long-run, though at different periods of time. In the short-run, only patience and interpersonal trust are (negatively) related to policies' stringency. These results are robust to the inclusion of some control variables, to changes in datasets or dependant variables' definitions, as well as to distinguishing between continents. Overall, our analysis identifies clear, highly-significant and robust correlation patterns using a rich dataset at the country level.

This dissertation is structured as follows: for each chapter the body of the chapter is first presented, followed by a dedicated appendix. Then, the dissertation proceeds to the next chapter until the general conclusion. Each chapter and each appendix contains a reference section exposing the related bibliography.

References

- Afriat, S. N. (1967). The construction of utility functions from expenditure data. International Economic Review, 8(1):67–77.
- Alfaro, L., Faia, E., Lamersdorf, N., and Saidi, F. (2020). Social interactions in pandemics: fear, altruism, and reciprocity. Technical report, National Bureau of Economic Research.
- Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. Econometrica, 70(2):737–753.
- Besley, T. and Coate, S. (1997). An economic model of representative democracy. The Quarterly Journal of Economics, 112(1):85–114.
- Brennan, G., Buchanan, J. M., et al. (1980). The power to tax: Analytic foundations of a fiscal constitution. Cambridge University Press.
- Breton, A. and Wintrobe, R. (1975). The equilibrium size of a budget-maximizing bureau: A note on niskanen’s theory of bureaucracy. Journal of Political Economy, 83(1):195–207.
- Buchanan, J. M. and Tullock, G. (1962). The Calculus of Consent, volume 12. Ann Arbor: University of Michigan Press.
- Choi, S., Fisman, R., Gale, D. M., and Kariv, S. (2007). Revealing preferences graphically: an old method gets a new tool kit. American Economic Review, 97(2):153–158.
- Coughlin, P. J. (1992). Probabilistic voting theory. Cambridge University Press.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. Journal of Economic Literature, 47(2):315–72.
- Downs, A. (1957). An economic theory of democracy. New York: Harper & Row.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. The Quarterly Journal of Economics, 133(4):1645–1692.

- Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. American Economic Review, 97(5):1858–1876.
- Galizzi, M. M. and Navarro-Martinez, D. (2019). On the external validity of social preference games: a systematic lab-field study. Management Science, 65(3):976–1002.
- Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., and others (eds.) (2020). World values survey: Round seven–country-pooled datafile version. Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. doi.org/10.14281/18241.13.
- Hale, T., Hale, A. J., Kira, B., Petherick, A., Phillips, T., Sridhar, D., Thompson, R., Webster, S., and Angrist, N. (2020). Global assessment of the relationship between government response measures and covid-19 deaths. medRxiv.
- Harrison, G. W. and List, J. A. (2004). Field experiments. Journal of Economic Literature, 42(4):1009–1055.
- Harstad, R. M. and Selten, R. (2016). Diminished-dimensional political economy. European Economic Review, 83:213–219.
- Heufer, J. and Hjertstrand, P. (2015). Consistent subsets: Computationally feasible methods to compute the houtman–maks-index. Economics Letters, 128:87–89.
- Heufer, J. and Hjertstrand, P. (2019). Homothetic efficiency: Theory and applications. Journal of Business & Economic Statistics, 37(2):235–247.
- Hotelling, H. (1929). Stability in competition. Economic Journal, 39(4):57.
- Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., and others (eds.) (2014). World values survey: Round six–country-pooled datafile version. <https://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>. Madrid: JD Systems Institute.
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives, 21(2):153–174.

GENERAL INTRODUCTION

- Lindbeck, A. and Weibull, J. W. (1987). Balanced-budget redistribution as the outcome of political competition. Public Choice, 52(3):273–297.
- Migué, J.-L., Belanger, G., Niskanen, W. A., Migué, J.-L., and Bélanger, G. (1974). Toward a general theory of managerial discretion [with comment and reply]. Public Choice, pages 27–51.
- Mueller, D. C. (2003). Public choice III. Cambridge University Press.
- Niskanen, W. A. (1971). Bureaucracy and representative government. Transaction Publishers.
- Persson, T. and Tabellini, G. (2002). Political economics: explaining economic policy. MIT press.
- Schnellenbach, J. and Schubert, C. (2015). Behavioral political economy: A survey. European Journal of Political Economy, 40:395–417.
- Smith, V. L. (1982). Microeconomic systems as an experimental science. The American Economic Review, 72(5):923–955.
- Thaler, R. F. and Sunstein, C. R. (2008). Nudge: Improving Decisions about Health, Wealth, and Happiness. Yale University Press.
- Varian, H. R. (1982). The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society, pages 945–973.
- Viscusi, W. K. and Gayer, T. (2015). Behavioral public choice: The behavioral paradox of government policy. Harvard Journal of Law & Public Policy, 38:973.
- Wilson, R. K. (2011). The contribution of behavioral economics to political science. Annual Review of Political Science, 14:201–223.
- Wittman, D. (1977). Candidates with policy preferences: A dynamic model. Journal of Economic Theory, 14(1):180–189.
- Zamir, E. and Sulitzeanu-Kenan, R. (2018). Explaining self-interested behavior of public-spirited policy makers. Public Administration Review, 78(4):579–592.

Chapter 1

Are Politicians Efficient Decision-Makers? A Revealed Preference Approach¹

Contents

1	Introduction	18
2	Related literature and behavioral conjectures	25
3	Experimental design	33
4	Revealed Preference Analysis	44
5	Results	48
6	Power analyses	57
7	Concluding remarks	62
	References	66

¹This chapter is an extended version of a working paper which is single-authored.

"[The private citizen] expends less disciplined effort on mastering a political problem than he expends on a game of bridge... Thus, the typical citizen drops down to a lower level of mental performance as soon as he enters the political field. He argues and analyzes in a way which he would readily recognize as infantile within the sphere of his private interests. He becomes a primitive again. His thinking becomes associative and affective."

Joseph Schumpeter, Capitalism, Socialism and Democracy, 1942

1 Introduction

In virtually all contemporary political regimes, only a fraction of citizens is responsible for implementing public policies which affect the entire population. An important scientific effort has thus been devoted to understanding how such individuals decide. In economics and political science, much of the scientific knowledge on policy-making is embedded into a rational choice framework. Under such framework, policy-makers are conceptualized as purposive, cool-headed individuals who optimize their behavior based on stable and transitive preferences in a resources-unconstrained and full-information environment (Mueller, 2003, Axelrod, 2015). However, surprisingly little is known about the empirical validity of such rationality assumption.² In the current study, we provide direct evidence on the relevance of the rational choice framework in explaining policy-makers' decisions based on an economic experiment with actual politicians.³

The relevance of the rational behavior hypothesis in the political arena has been questioned by many scholars for a long time. Indeed, several early thinkers (such as Adam Smith, John Stuart Mill or Joseph Schumpeter) already argued that the

²There exists many different definitions of "rationality", which we will not attempt to summarize here. The definition which we adopt is the one which is commonly employed in economics and which can be expressed as: "a decision is only rational if it is supported by the best reasons and achieves the best possible outcome in terms of all the goals" (Simon, 1995). In particular, such a definition entails that individuals are capable of clearly identifying their goals (or "preferences") in any context and to make decisions that are consistent with them.

³For the sake of repetition avoidance, we will take policy-makers and politicians as synonyms throughout this text. Of course, it is not to say that policy-makers are necessarily politicians, as many people who were not elected ultimately participate actively in policy-making (starting with governments' staffs).

rationality principle that governs individuals in the marketplace is much less likely to stand in political life (see exposure and references in Schnellenbach and Schubert, 2015). Building on similar arguments and highlighting some key results from economic experiments, several scholars called for moving from the standard utility maximization framework to a behavioral approach of political decisions (e.g., Simon, 1995, Jones, 2001, Kliemt, 2005, Wilson, 2011).

Recently, many different strands of the literature have started to incorporate behavioral insights into the study of political behavior. However, the bulk of studies focus on the demand side of the market for policies (voters). In particular, an important part of the literature has focused on changes following the introduction of citizens' choice anomalies (DellaVigna, 2009, Harstad and Selten, 2016, Palfrey, 2016)⁴ or on the issue of maximizing welfare from a behavioral perspective (Bernheim and Rangel, 2007, DellaVigna, 2009, Mullainathan et al., 2012).⁵ Yet, all these studies implicitly maintain the assumption that supply-side actors (policy-makers) will rationally react to citizens' biases or behaviorally-grounded policy recommendations. Holding together that people are biased but that the people who will correct the biases are not appears blatantly paradoxical (Berggren, 2012, Viscusi and Gayer, 2015). Moreover, when scholars do incorporate behavioral insights to focus on politicians' behavior, they essentially project behavior observed with non-politician pools (mostly from student samples) onto political decisions without further evidence that such a projection accurately describes politicians' behavior (Wilson, 2011, Schnellenbach and Schubert, 2015, Palfrey, 2016). Indeed, in order to coherently derive useful insights from studies advancing evidence based on non-politician samples, one

⁴For instance, politicians may schedule implemented policies by taking into account voters' limited attention (e.g. discussing important reforms during the summer break or during the Olympics).

⁵For instance, the likelihood of people's irrational behavior due to cognitive limitations has been perceived as legitimizing a new kind of public intervention designed to correct individual biases in order to improve social welfare. However, relying on heuristics or habits to make a decision may be welfare-enhancing, in particular in the presence of standard market failure when strictly rational behavior from individuals is not collectively optimal. Choosing when to launch a debiasing attempt is therefore a public policy dilemma.

must determine whether politicians as individuals markedly differ or not in the way they behave from non-politicians (especially students).⁶

At the same time, the decision-making ability of politicians appears crucial for an efficient implementation of public policies. Indeed, if politicians behave irrationally, then it is likely that the resulting policy outcomes would be sub-optimal and would waste public funds. Furthermore, the incentive schemes and institutions surrounding policy-makers' behavior have an impact on individual behavior, and the underlying assumption about how these people would react to these structures is consequently important. If incentives assume rational responses from politicians, then cognitive limitations may reduce and even reverse their effectiveness. For instance, Cooper and Kovacic (2012) show in a behavioral model of a regulator how the combination of short-term incentives and common heuristics or myopia results in output-based rather than outcome-based policy choices, hence generating inefficient policies. Similarly, Viscusi and Gayer (2015) provide a list of biases and anomalies on the part of regulation agencies along with several real-life illustrative examples to underline the usefulness of the behavioral paradigm in explaining political actors' behavior.⁷ Assessing whether the rational choice framework is a reasonable approximation for politicians' behavior at the individual level is therefore of general interest.

In our views, there are several explanations for why political leaders may differ from the remainder of the population in adopting a rational behavior.⁸

First, politicians operate in a highly complex environment: they are exposed to many complicated pieces of information of various nature, their decisions are often multi-dimensional, and the resulting policies produce uncertain results. They

⁶Obviously, the advantage of non-politician samples, and particularly student samples, is their relative accessibility and cheapness. On the contrary, collecting information on political behavior from politicians or policy-relevant issues based exclusively on politician samples would be very difficult and expensive. However, by knowing how politician and non-politician samples compare on general issues (say sensitivity to some famous cognitive biases, preferences over risky or social domains, and so on), one could be able to exploit the observed differences when assessing politicians' behavior in a given context by extrapolating from non-politician studies.

⁷Alternative examples can be found in Jones (2001), Rachlinski and Farina (2001), Berggren (2012) or Schnellenbach and Schubert (2015).

⁸Some of these arguments have been precisely mentioned by scholars to advance the plausibility of the rational choice framework in studying politicians' decisions. Yet, as we argue for each of the below-mentioned mechanisms, the direction of the effect is not *a priori* straightforward: there are also reasons to believe that politicians may act less rationally than the remainder of the population.

are thus likely to depart from rationality due to the complexity of their decision environment (e.g., using heuristics, see Gilovich et al., 2002).⁹ At the same time, political elites often have access to several resources (e.g., data, staff, experts) which may help them reach accurate decisions even in such complex environment (DellaVigna, 2009, Jacobs, 2011). In addition, politicians accumulate experience in political decision-making which may render them more efficient in such decisions and reduce behavioral anomalies (Fréchette, 2016).

Second, the political decision-making environment is atypical insofar as policy-makers are being held publicly accountable for their actions, with possibly important impacts on political careers when they behave irresponsibly (Ashworth, 2012). Relatedly, politicians are daily exposed to stiff competition for votes, which may discipline their actions (Wallerstein, 2004). In other words, the incentives are high in political action, which could spur politicians to spend more efforts on reaching an accurate judgement and achieving a higher level of sophistication (DellaVigna, 2009). Such reasoning would indeed match the standard median voter model (Downs, 1957) where politicians are expected to implement voter's wishes to the best of their ability. However, public screening and competition may also pressure policy-makers into adopting an inefficient behavior, such as (dis)favoring the status-quo or making emotionally-driven decision (Jennings, 2011, Bischoff and Siemers, 2013).

Third, political candidates (hence ultimately politicians) are not sorted at random from the population: there is a highly-influential political selection process at play (Osborne and Slivinski, 1996, Besley and Coate, 1997). Several studies indeed find that political candidates systematically differ in some physical or psychological traits as compared to the remainder of the population (Caprara et al., 2010, Besley and Reynal-Querol, 2011, Dietrich et al., 2012, Muller and Page, 2016, Schreiber, 2017). In particular, there exists some evidence that policy-makers display different cognitive abilities (Dal Bó et al., 2017, Chaudhuri et al., 2020), which have been found to relate to the propensity to behave consistently (Chen et al., 2013, Choi et al., 2014, Cueva et al., 2016, Amador-Hidalgo et al., 2021). Moreover, that in-

⁹We acknowledge that using heuristics may ultimately result in an efficient use of limited cognitive resources, so that interpreting it as a deviation from rationality is debatable. However, as argued by Schnellenbach and Schubert (2015), the use of heuristics on the part of politicians may not result from the willingness to overcome one's own limitations but may relate to political calculus. For instance, a politician may abide by a common rule-of-thumb (say, immigration is bad for employment) in order to signal himself on the ideological scale to his electorate, even though he may not be fully ready to endorse it once in office.

dividuals with high decision-making abilities are selected as politicians appears a desirable feature. Voters have indeed been found to value competence among their representatives (see Dal Bó and Finan, 2018, for a survey).

In the present study, we apply the revealed preference methodology to assess the consistency of observed decisions with utility maximization on the part of a sample of politicians. To the best of our knowledge, we are the first study to focus on the consistency of politicians' behavior with the revealed preference axioms at the individual level.¹⁰ Specifically, we recruit volunteers within the population of French local politicians (henceforth "Politicians") to participate in an *induced budget experiment*.¹¹ Each subject is exposed to a large series of decisions asking him to divide an amount of money between himself and another subject from different linear budgets (Andreoni and Miller, 2002, Fisman et al., 2007).¹² Treating both subjects as consumer goods, we then test each individual dataset's compatibility with the Generalized Axiom of Revealed Preference (GARP) and measure the extent by which the dataset departs from the axiom using several rationality indices (Afriat, 1972, 1973, Houtman and Maks, 1985, Varian, 1990, 1993). Indeed, consistency with GARP is equivalent to assuming that some well-behaved utility function exists that rationalizes the data (Afriat, 1967, Varian, 1982). We then compare Politicians' choice consistency with the one observed from two control groups: a convenience student control group (henceforth "Student Control", SC) and a control group taken from the general French population (henceforth "Non-Student Control", NSC).¹³ In order to reduce possible confounds and facilitate statistical comparisons, the NSC

¹⁰In the remainder of this study, we will consider "rational behavior", "choices' consistency" and "utility maximization" as synonyms, a semantical practice that is implied by the revealed preference methodology.

¹¹The term "induced budget experiment" was coined by Banerjee and Murphy (2011) to refer to experiments designed to elicit subjects' valuation of some good. Such class of experiments differs from the class of "induced value experiments" where the objective is to control values (say demand and supply) in order to evaluate the performance of an allocation mechanism or institution.

¹²Focusing on a distributive setting is legitimate considering that most of politicians' daily activity consists in allocating some fixed budget among a series of possible uses, hence across competing groups of citizens. In other words, politicians are constantly trying to determine which groups are to be favored and which are not.

¹³Throughout this article, we abide by the following convention: capital letters are reserved for experimental groups. For instance, "politicians" refers to the political elite in general whereas "Politicians" refers to the specific individuals who participated in our experiment. The same applies to Students and Non-Students.

was constructed so as to resemble Politicians based on observables (gender, age and occupation). Finally, our sample of politicians is composed of two sub-groups: one group came into the laboratory and followed standard procedures (henceforth "Politicians from the Lab") whereas the laboratory came to the other group which participated in a "field" version of the experiment (henceforth "Politicians from the Field"). Comparing both sub-groups enables to question the validity of a laboratory experiment in accurately describing the behavior of such a very specific pool of subjects.

We focus on local-level politicians for several reasons.

First, following the decentralization movement that happened in many western democracies during the 80s, the share of public goods and services that are locally handled has tendentially increased over time. In 2019, the local public spending in France amounted to € 271.1 billions, which accounted for approximately 20% of the total public spending and 11% of GDP (INSEE data). In other words, an important dimension of citizens' daily economic life is managed by local public decision-makers (e.g. management and maintenance of towns' streets and cemeteries, treatment of domestic waste, management of towns' general sanitation, supply of funeral homes, primary schooling, leisure and sport centers, libraries, green areas, fire stations and so forth).

Second, considering that political careers often start at the local level, focusing on this very first rung of the political ladder is valuable as it may determine observed behavior at higher layers. More generally, several studies emphasize that political entry at the local level is an important driver of well-functioning democratic regimes (e.g., the prevalence of clientelistic politics, lower political competition, efficiency in public good provision or governance, etc. Martinez-Bravo, 2014, 2017, Martinez-Bravo et al., 2017).

Third, although the empirical evidence is still scant, local politicians appear socio-demographically closer to their constituents than upper-tiers politicians and display a greater level of heterogeneity in observables (see Gulzar, 2021, and references therein). In particular, they are generally not perceived as "elitist" decision-makers with a very specific economic and social position in society prior to their election. Nor had they received a specific political education or training in most cases. Building a valid control group taken from the population is thus arguably less likely to be riddled with biases in unobservables.

Fourth, local politicians are logistically less complicated to recruit than higher-layers politicians due to their large number, their globally less-constrained schedule, and their important geographical dispersion among the French territory.

Globally, our study relies on well-known revealed preference techniques allowing to expose a unique subject pool to a theory-grounded non-parametric test of individual-level rational behavior in a tightly-controlled incentivized decision environment. Overall, we find that the majority of individual datasets include some choice inconsistencies, but such inconsistencies are neither quantitatively nor qualitatively sizeable on average. The utility maximization model thus seems a rather useful approximation for the analysis of actual policy-makers' decisions. Furthermore, we uncover only weak evidence that Politicians differ from a comparable control group: no matter the rationality index under scrutiny, we find that Politicians are slightly closer to GARP consistency than the control group but also that most differences are not statistically significant. Consequently, politicians optimize behavior at least as efficiently as non-politicians. However, the superior decision-making ability that is generally implicitly assumed in the literature on the part of politicians actually seems only marginal at the individual level. On the contrary, we find strong support in favor of Politicians being farther away from rational choice than a convenience student sample. Such a result has important implications with respect to the generalizability of experimental results obtained with convenience samples in the study of political decisions. Furthermore, we find little evidence of any difference among politicians between our lab and field sessions in terms of rationality, which is an important result for future experimental studies relying on a similar subject pool.

The remainder of the chapter proceeds as follows. Section 2 relates the current study to different strands of literature and underlines some behavioral conjectures. Section 3 depicts the experiment in details. Section 4 specifies the definition of rationality adopted here and presents several rationality measures. Section 5 exposes the main results based on between-groups comparisons. Section 6 presents a power analysis to check the statistical robustness of our results. Finally, section 7 proposes a discussion and concludes.

2 Related literature and behavioral conjectures

Our study mainly connects to three strands of the literature. First, some local public finance studies rely on the revealed preference methodology to assess the relevance of existing models of governments' behavior based on observational data. Second, the literature on political selection emphasizes that ability and competence are important drivers of political candidacy. Third, a rapidly growing number of experiments, mainly in political science, are testing deviations from rational choice with actual politicians. For each strand of literature, we summarize the results and emphasize the extent by which our study relates to it. Based on such a literature review, we derive several behavioral conjectures as to how our subjects might behave during the experiment.

There have been some attempts within the realm of local public finance to evaluate the compatibility of existing economic models of political decision-making at the local level by applying the revealed preference methodology to observational data.¹⁴ For instance, Turnbull and Chang (1998) evaluate the compatibility of American municipalities' spending pattern with the median voter model by implementing a GARP test. They find that in most cases the data is consistent with the utility maximization hypothesis. A similar conclusion is derived for French municipalities by Baudry et al. (2002). An alternative public sector model based on bureaucratic behavior is tested by Chang and Turnbull (2002) with a similar methodology. They find that Taiwanese local governments are mostly compatible with bureaucratic behavior aiming at maximizing public spending. Finally, other studies focus on the possible factors explaining deviations from GARP on the part of municipalities, in particular institutions and demographics (Turnbull and Geon, 2006, Salvino et al., 2012).

However, such studies do not directly test for the compatibility of individual

¹⁴Another strand of literature builds upon a theoretical framework that assumes a rational decision-maker over multiple time periods to derive testable hypotheses for parameters from a simple regression model. The model is then applied to local governments' data. If the null hypotheses of rational behavior cannot be rejected, the study generally conclude that the rational choice framework is consistent with the data. Contrary to these approaches, we do not make any parametric assumptions to test for rationality, which also enables to avoid econometric issues. See for instance Holtz-Eakin and Rosen (1989) or Borge and Tovmo (2009).

politicians' behavior with revealed preference axioms. Inferring the rationality of the decision-makers from such studies would indeed be misleading for several reasons.

First, actual political decisions result from a two-folded collective process: they rely on voting rules within the local board and they involve a possibly large series of actors (politicians, public servants, regulators, lobbies, citizens, etc.). Assessing whose preferences are observed from observational data is thus unclear. Furthermore, it is ultimately the objective of existing studies implementing the GARP test on observational data to test whether observed patterns are compatible with existing models of behavior (mainly the median voter model where politicians basically implement voters' preferences). Such an issue connects to the very large economic literature on modelling public decision-making (see Mueller, 2003, for a review).

Second, such observational studies are not testing the compatibility of the data with some utility function (as would a non-parametric test do). Rather, these tests rely on a semi-parametric procedure that incorporates some functional elements derived from the underlying utility model whose compatibility with the data is scrutinized. On the contrary, we do not seek to assess the validity of one specific model of politicians' behavior. Our objective is to determine whether politicians as individuals can be modelled as utility maximizers in general, and whether they differ from other groups in this regard.

Third politicians' decisions are often made in a very peculiar environment which is likely to influence their decision-making ability, due for instance to its complexity. Such a complexity may favor the emergence of specific institutional settings which supply simple conceptions of political issues along with ready-made solutions (or "tunnel visions", see Viscusi and Gayer, 2015). These institutions enable the persistence of inefficient policies (e.g., piling-up reforms) in contrast with a comprehensive approach that would identify and reduce inefficiencies.¹⁵ Consequently the observed policies are not directly informative as to the politicians' ability to make consistent choices as compared to other people in the same decision environment. Besides, these elements of the decision environment are also likely unevenly distributed across jurisdictions, which makes an aggregate analysis even harder to interpret.

¹⁵At the same time, the existence of those institutions is also a response to the complexity of the decision environment, which ultimately makes them "rational" in the sense that they facilitate the decision-making process.

Fourth, and more generally, observational data may even distort the empirical conclusions. Indeed, bad policies may result from self-interested maximizers who select the perfect policy mix of good policies and bad, yet electorally-relevant, policies (Schnellenbach and Schubert, 2015). For instance, Dal Bó and Di Tella (2003) show how even talented and honest politicians may be forced by pressure groups to implement bad policies, which in turn may enhance the prevalence of "inept" policy-makers. Similarly, Dal Bó et al. (2018) demonstrate how bad policies may also result from voters' cognitive limitations (especially anticipation failure), while Gustafsson (2019) emphasizes that politicians may be incentivized to implement even inefficient policies as a signalling device for competence. Ultimately, we do not know whether what we conceive as being irrational (the policies) stem from politicians being behavioral human beings (hence susceptible of failures) or because there exists some "public choice incentives" that guide them towards their choices (Viscusi and Gayer, 2015). Ultimately, disentangling the effect of politicians' individual irrationality from other possible effects based on the outcome of political decision-making is therefore not trivial. On the contrary, using a laboratory experiment allows to finely control the decision environment and to propose a test of rational choice based on a simple distributional task that moreover accounts for individual heterogeneity in behavior.

Our study also relates to the political selection literature which aims at understanding the forces surrounding the quality of political leaders. Part of this literature is indeed focused on politicians' ability or competence, to which rationality can be seen as a subset. Early theoretical models underline various comparative advantages for the less able citizens to enter politics, which could supply a straightforward hypothesis regarding policy-makers' rationality (Caselli and Morelli, 2004, Messner and Polborn, 2004, Dal Bó et al., 2006, Mattozzi and Merlo, 2008). One basic mechanism is the following: poorly-skilled citizens have few opportunities in the private sector to obtain a high wage as compared to highly-skilled citizens, which results in a leakage of highly-skilled citizens and a concentration of poorly-skilled citizens into politics. However, ability is not the only individual characteristic that is likely to matter for political selection. In particular, other models emphasize the importance of honesty or public-service motivations in determining the supply of political can-

didates (Besley, 2005, Dal Bó and Finan, 2018).¹⁶ If such characteristics prevail in a given setting, then we may not observe any difference between politicians and the remainder of the population as far as rationality is concerned.

Regarding the empirical evidence on the relationship between political selection and ability/competence, most studies focus on proxies such as education or income due to data shortage. Overall, the evidence suggests that politicians fare better than the remainder of the population (see the survey in Gulzar, 2021). However, such proxies may unfortunately be correlated with other characteristics (like social class) and fail to finely account for competence or ability.¹⁷ Another way to investigate such an issue that goes beyond the sociodemographic differences between policy-makers and citizens is proposed by Besley et al. (2017): they estimate innate ability based on the residuals of a Mincer equation using Swedish census data and find that such a variable correlate with cognitive and leadership ability, as well as with measures of political and policy success (such as voters' support, reelection likelihood or citizens' satisfaction). Other studies attempt to directly measure cognitive abilities of politicians. For instance, Dal Bó et al. (2017) rely on a large dataset combining Swedish politicians and a representative sample from the Swedish population and find that the former score higher than the latter in measures for both cognitive abilities and leadership. A contrasted result, more consistent with the early theoretical perspectives, is however obtained by Chaudhuri et al. (2020) who expose local politicians and citizens in India to the Raven's Progressive Matrices test and find that the results are sensitive to political experience: inexperienced politicians display lower cognitive abilities than both experienced politicians and citizens.

As a consequence, the existing evidence from this literature suggests that politicians may be selected based on ability, which could translate into a higher propensity to make consistent choices. However, all studies (either theoretical or empirical) do not systematically point towards a positive relationship between the selection into politics and measures of ability. One likely possibility would indeed be that such a relationship relies upon the prevailing institutions surrounding the selection process. For instance, if corruption is rampant in a given country, for instance in a new democracy, then it likely affects the quality of the politicians pool and may lead to

¹⁶We come back to this specific literature with greater details in the chapter 2.

¹⁷Two exceptions are Dal Bó et al. (2017) and Pedersen and Dahlgaard (2021) who compare politicians to their siblings to control for family characteristics and find the same results as the remainder of the literature (higher education and income).

the selection of low-quality candidates (see the survey in Dal Bó and Finan, 2018). Besides, the link between ability and rationality is not necessarily straightforward, whereas we propose to directly gauge politicians' consistency in their decisions and to compare it with the behavior observed in a comparable population.

Finally, our study is particularly connected to the growing literature questioning the rationality of political elites at the individual level. Specifically, several recent studies in experimental political science and public administration use extensive surveys as well as large-scale experiments to study common choice anomalies and cognitive biases using a pool of actual politicians. For instance, some studies evaluate the relevance of prospect theory as opposed to the standard rational choice benchmark in explaining actual policy-makers' decisions in a risky environment. Both Fatas et al. (2007) and Linde and Vis (2017) find that politicians deviate from expected utility theory in the way predicted by prospect theory but to a lesser extent than non-politicians. However, both studies rely at least partly on hypothetical decisions and involve relatively small samples. Using a large incentivized experiment in three democratic countries, Sheffer et al. (2018) find that political leaders are actually more subject to some choice anomalies than the respective populations (in particular a higher propensity to escalate commitment when facing sunk costs and a higher sensitivity to the status-quo bias).

Politicians also appear subject to a diversity of framing effects (Linde and Vis, 2017, Walgrave et al., 2018, Arceneaux et al., 2018, Sheffer and Loewen, 2019, Baekgaard et al., 2019a), and are sometimes more sensitive to such effects than citizens (Sheffer et al., 2018). There also exists evidence that politicians rely on heuristics in formulating a judgment or making a decision (Vis, 2019, Stolwijk and Vis, 2020).¹⁸ However, elite decision-makers in general also appear more efficient at using such heuristics when processing information than the remainder of the population (Hafner-Burton et al., 2013). Relatedly, Fisman et al. (2015b) find that American

¹⁸Relatedly, some scholars have stressed the empirical relevance of modelling politicians' behavior as heuristic-based. See for instance the study by Le Maux (2009) on the hill-climbing heuristic which he tests using observational data on French departments' expenditures.

elites¹⁹ are closer to rational behavior than other groups. Similarly, LeVeck et al. (2014) find that international elites exhibit higher levels of strategic reasoning than a convenience student sample. Several survey studies also emphasize that politicians are prone to engage in various forms of motivated reasoning (Nielsen and Moynihan, 2017, Christensen et al., 2018, Schönhage and Geys, 2021)²⁰, possibly more than the remainder of the population (Baekgaard et al., 2019b, Esaiasson and Öhberg, 2020), and even to resist debiasing attempts (Baekgaard et al., 2019b, Christensen and Moynihan, 2020). Finally, recent studies scrutinize the compatibility of observed Mayor’s behavior with a rational processing of new information. For instance, resorting to a very large experiment with Brazilian Mayors to assess the impact of research information on their beliefs and behavior, Hjort et al. (2021) observe few deviations from Bayesian learning (in particular no confirmation bias, no beliefs’ reinforcement and no politically-motivated updating). Similarly, Lee (2021) presents American local and state policy-makers with expert findings and observes belief updating in the sense of the presented evidence.

The current study primarily differs from the above studies in the definition of ”rationality” that we adopt. We are not addressing the way politicians process the information or endorse some specific cognitive limitations. Instead, our focus lies on the degree of internal consistency of observed individual-level decisions made by politicians, hence on their ability to behave as if they had coherent preferences and were maximizing some utility function. As such, our study provides a direct non-parametric test of the utility maximization hypothesis instead of a test of some well-known deviations from it.²¹ Furthermore, acknowledging that politicians may endorse common cognitive limitations that are observed with other pools at a varying degree does not provide any information as to the overall importance of such biases

¹⁹The study by Fisman et al. (2015b) relies on almost the same design as the one we use. However, the definition of elites adopted in their paper is not restricted to political elites. Indeed, their results rely either on future elites (students from Yale), with no certainty that they will enter the political arena, or on actual elites from the general American population as defined by thresholds on education and income, hence not necessarily politicians. A similar criticism applies to Hafner-Burton et al. (2013).

²⁰Motivated reasoning refers to the fact that people process information in a way that serves their interest, hence not necessarily in a way that maximizes predictive accuracy.

²¹The reasons behind departures from rational behavior that we may observe may be linked to cognitive biases of various forms, some of which having been scrutinized by the above literature. Attempting to discriminate between explanations behind deviations from GARP is however beyond the scope of this chapter.

for actual decisions. On the contrary, our study is based on continuous measures of rationality which quantify the efficiency loss implied by deviating from GARP. Finally, the number of incentivized experiments among existing studies is limited (often by the study's topic which does not easily lend itself to incentivizing participants), they generally do not include a control group, and when they do they focus on the remainder of the population and not specifically on a comparable group. On the contrary, our study resorts to an incentivized measure of behavior which we use to compare a sample of politicians with a sample selected from the general population to resemble the latter based on simple observables.

To summarize, existing evidence shows that in many respect politicians fail to adopt a rational behavior. We thus expect at least some politicians to be inconsistent in their decisions, which translates into the following conjecture:

Hypothesis 1 (utility maximization from politicians, H^1): Politicians fail to comply with GARP in at least some extent.

However, the evidence concerning the proximity with rational behavior among political elites as opposed to a comparable population is mixed. Such a comparison is indeed likely to depend upon both the definition of rationality and the definition of "political elite" that is adopted. Furthermore, most studies do not actually compare politicians with other people. They rather focus on assessing whether deviations from rationality that is observed with non-politician populations is also observed with politicians.

At the same time, studies applying the revealed preference methodology to experimental subjects from representative samples find that most subjects violate GARP but that average behavior is quite close to optimizing behavior (Choi et al., 2014, Fisman et al., 2015b, 2017). Such studies also reveal sizeable heterogeneity in rationality measures depending on individual characteristics. We therefore do not make a strong statement as to the direction of a possible effect when comparing the rationality of Politicians and Non-Students. Moreover, these studies underline that student samples as well as young people display higher rationality scores than other fringes of the population. As a consequence, we consider two behavioral conjectures for group comparisons based on the existing literature:

Hypothesis 2 (group comparison in utility maximization, H^2): Politicians do not significantly differ from a comparable non-politician control group in adopting rational behavior.

Hypothesis 3 (group comparison in utility maximization, H^3): Students display higher choice consistency than both other groups.

Finally, we do not make a strong statement as regards a possible difference between Politicians from the Lab and Politicians from the Field. From a theoretical perspective, we could expect the costs to come into the laboratory to be higher than the costs to participate in field sessions. It was indeed one objective of the field sessions to facilitate participation by reducing both economic (e.g., travel costs) and psychological costs (e.g., not knowing the environment). Our ultimate goal was to complement the sample we obtained in the Lab, in order to achieve a greater diversity of politicians' individual profiles. Given that experimental subjects self-select into participation based on a cost/benefit analysis (Slonim et al., 2013)²², changing the parameter of this analysis may change the selection process and attract different fringes of the overall politicians' population.²³ Such a change may affect our results on decision-making ability. For instance, Slonim et al. (2013) find student participants to be more consistent in saving decisions and to score higher in the Cognitive Reflection Test than student non-participants (the latter finding being replicated in Thiemann et al., 2022). The authors indeed argue that experiments attract individuals who like to be intellectually challenged. Along these lines, we may hypothesize that Politicians from the Lab behave more rationally than Politicians from the Field. However, the cost dimension is not the only one to consider when comparing the Field and the Lab. Indeed, the recruitment process in the Lab partly relied on word-of-mouth, which introduces some dependance structure across

²²Several studies emphasize that participants in economic experiments differ from non-participants in many different ways, among which income, majoring in economics/business, likelihood to volunteer, having an interest in the research domain and various types of preferences (e.g., Cleave et al., 2013, Slonim et al., 2013, Frijters et al., 2015, Thiemann et al., 2022).

²³We note that running the same experiment in the field does not eliminate the selection process: in any case subjects decided to participate in an experiment, which is a decision that is likely to generate some self-selection. However, participating in the laboratory is arguably more demanding than participating in the field, which is consistent with a slightly higher number of participants per session in the Field than in the Lab.

sessions. On the contrary, the field sessions were totally independent from one another. Such a difference in the recruitment procedures may have an impact on the results, through for instance a shared knowledge of the content of the experiment between early participants and potential participants. This could contribute to the selection of some specific subjects but may also directly affect the results (e.g., if participants have the time to think about what they would do in advance). In addition, Politicians from the Field had a higher probability than lab participants to have already been acquainted with other participants in the same session because they took part in the same intermunicipal government. Although anonymity was strictly guaranteed throughout the experiment, there is a possibility that such a feature interacted with observed behavior anyway.²⁴

3 Experimental design

3.1 Subject pools and procedures

In our experiment, we compare the behavior of subjects from one group of interest (Politicians) with the behavior of subjects from two control groups (Student Control and Non-Student Control). Our sample of politicians is composed of two sub-groups: Politicians from the Lab and Politicians from the Field. These sub-groups aim at implementing complementary recruitment procedures in order to obtain a sufficiently diverse sample of Politicians. Overall, 353 people volunteered to participate in the experiment, among which 139 Politicians (67 from the Lab and 72 from the Field), 123 Students and 91 Non-students, divided into 32 experimental sessions.²⁵ These sessions were structured into two waves, each comprising two groups: Politicians from the Lab and Students participated from late 2018 until early 2019 (first wave) whereas Politicians from the Field and Non-Students partici-

²⁴We also note that the impact of priori acquaintance is likely limited due to the fact that the invitation to participate in the Field stressed that all local politicians were invited to participate, not only intermunicipal councillors. As a consequence, many Politicians from the Field participated in the experiment without regularly participating in intermunicipal meetings.

²⁵For a detailed listing of all sessions, see table A.1 in appendix A.1. Sessions' dates and times were accommodated in order to propose a wide range of slots to the atypical populations we are interested in. Specifically, some sessions were organized in the evening, during the university's holiday and even on Saturdays.

pated from late 2019 until early 2020 (second wave).²⁶ All sessions happened either at the LABEX-EM of the University of Rennes I in France (lab sessions) or using a "mobile lab" with tablet computers in several towns from France's Brittany region (field sessions).

The Politicians from the Lab were recruited thanks to the support of the Ille-et-Vilaine branch of the Association des Maires de France (AMF). The AMF is the biggest association of french Mayors. Its membership does not rely on any specific requirements (e.g., city's characteristics or political affiliation) and it relies on a departmental division, which greatly facilitates a geographically-oriented recruitment. The association agreed upon conveying our participation call and sharing a large mailing list which enabled us to contact the bulk of members of local councils in Ille-et-Vilaine (a department within the western Brittany Region). Interested politicians could then register by e-mail to one of the proposed lab sessions and come to the laboratory.

The Politicians from the Field were recruited thanks to the support of 7 inter-municipal governments in both Ille-et-Vilaine and Finistère (another department from the same region of Brittany). These intermunicipal governments agreed upon hosting a tablet version of the experiment directly within their facilities at dates and times when intermunicipal meetings were usually held. The field sessions were organized at the intermunicipal level and not at the strictly local level for two main reasons. First, it provided an access to very heterogeneous profiles at once, enabling us to avoid choosing some towns rather than others or to risk that one session falls short of volunteers. Second, intermunicipal governments benefited from more important logistical resources to support the experiment than most towns. The choice of intermunicipal governments was driven by the following process: we focused on the intermunicipal governments around Rennes and Brest (which are the two largest cities in Ille-et-Vilaine and Finistère), mainly for logistical reasons, and we ultimately organized a session in every intermunicipal governments where its President approved the project.²⁷ In order to foster participation, the practical

²⁶It is worth noting that the last session occurred the 02/01/2020, hence approximately a month before the COVID-19 was recognized as a public health issue in France. As a consequence, our experimental measures are not biased by the epidemic.

²⁷There is one exception to this rule: we chose not run the experiment in Brest due to a too small number of volunteers (only two).

implementation of the recruitment procedure was delegated to each intermunicipal government.²⁸ Politicians from the Field thus participated to the experiment in a much more familiar environment than Politicians from the Lab.

It should be noted that we did not focus on Mayors only, neither in the lab nor in the field. Indeed, decisions in local towns do not solely rely on the Mayors' wishes but are rather collective decisions within the local council implying some kind of voting mechanism. Among the politicians who ultimately participated, 10 Mayors, 28 Vice-Mayors and 28 local councillors (among whom 2 were former councillors) and one former regional councillor came into the lab, and 21 Mayors, 30 Vice-Mayors and 21 local councillors came to the field sessions. Besides, 30 lab subjects (44.78%) and 45 field subjects (62.5%) also held an intermunicipal term by the time of the experiment.²⁹ Besides, though the sample size as well as the geographic concentration of the data preclude any claim to representativeness, we acknowledge a certain diversity regarding Politicians' background. A total of 89 different towns are indeed represented in our sample, with population ranging from 341 to 222,104 inhabitants (median 3,966) and income per inhabitant ranging from 10,873€ to 19,291€ (median 13,930€) in 2019.

The Students were recruited through ORSEE (Greiner, 2015) within a pool of

²⁸Specifically, we wrote the invitations in agreement with the intermunicipal governments, and then the secretariats handled both sending the invitations and managing the volunteers until the experiment.

²⁹The difference in proportions of Mayors/local councillors and intermunicipal councillors between the lab and the field could be indicative of some participation bias. Yet another interpretation would emphasize that field sessions happened precisely within intermunicipal governments. As such, and though the participation call specifically pinpointed that any local politician could come in both cases, we expected a higher turnout among intermunicipal councillors than among the remainder of local politicians. This could also explain the status discrepancy, considering that Mayors are more likely to hold an intermunicipal term. This is especially true for the 2014 election, because of a change in the election rules at the intermunicipal level. Prior to 2013, there were separate election systems between for municipalities and intermunicipalities. Such a system was replaced with a signposting system in 2013, so that for all french towns the list at the intermunicipal level is based on the list at the municipal level. More specifically, both lists are the same for towns below 1000 inhabitants and the seats at the intermunicipal council are attributed depending on the rank on the municipal list (starting with the candidate for mayor, then the candidates for vice-mayors and then the candidates for local councillors). For towns above 1000 inhabitants, the intermunicipal list constitutes a subset of the municipal list but the first quarter must be identical and the rank order must match across the two lists (e.g., a candidate for vice-mayor cannot be ranked below a candidate for local councillor). As a consequence, the likelihood of having mayors as intermunicipal councillors is high under such system.

Table 1.1: Summary of experimental groups

Group	Sessions	Recruitment	<i>N</i>
Politicians (Pol)	Laboratory	Association des Maires de France (AMF)	67
Politicians (Pol)	Field	Brittany's intermunicipal governments	72
Non-student Control (NSC)	Laboratory	Ads, social media, ORSEE, posters, etc.	91
Student Control (SC)	Laboratory	ORSEE	123

students studying in Rennes, mostly undergraduates in economics at the University of Rennes I. The Non-Students were recruited through several channels: announcement in a popular regional newspaper (Ouest France), advertisement through students' mailing lists (administrative and ORSEE) and through the university/ laboratory's social networks, as well as posters in local shops, markets, community centers, leisure associations and public spaces in Rennes. Interested people were asked to fill in an online form with basic socio-demographic questions (gender, age and occupation). The answers to such questionnaire were used to summon volunteers so that the proportions in these variables match as much as possible between Non-Students and Politicians (full sample). The objective was indeed to compare Politicians with a similar control group, except for the elected status. The table 1.1 summarizes the information on the experimental groups.

All sessions followed the same procedures: upon arrival subjects were greeted and explained the most basic rules prevailing during the experiment, a ball-in-the-urn process randomly attributed them stations in the laboratory, they listened and read the instructions all together, they played the induced budget experiment programmed in *z-Tree* (Fischbacher, 2007), they answered a short self-declared socio-

demographic questionnaire, and finally they individually received their paycheck.³⁰ Instructions had been recorded before the experiment took place, which guarantees that all participants face virtually identical rules across sessions and knew about it within sessions. Upon receiving their paycheck, subjects were asked to sign a non-disclosure form aiming at reducing the risk of contagion across sessions (in particular for Politicians from the Lab and Non-Students).³¹ The game lasted 45 minutes on average and all amounts were expressed in Experimental Currency Units, with $100 \text{ ECU} = 7\text{€}$.

During the game, each subject made 30 decisions. For each decision, each subject is randomly matched with another participant in the room. In a given pair of subjects, both subjects makes a decision for which there is one decision-maker and one passive subject. In other words, each subject is both a decision-maker in her own decision and a passive subject in the decision of the other subject. Subjects change partners across decisions (stranger matching). No feedback was implemented across decisions so that recipients did not know the results of the dictators' decisions at any point in the game. More generally, the instructions specifically emphasized that participation was strictly anonymous and that no participant would obtain information about the other participants' decisions or payoffs during the experiment. At the end of the experiment, each subject's payoff for the entire game was computed as the sum of i) the ECU she kept from one randomly-selected decision among the decisions she made, and ii) the ECU she received from one randomly-selected decision among the decisions other subjects made for her when she is a passive subject. Payoffs were individually displayed only at the end of the experiment (prior to the questionnaire). Each subject's payoff was composed of a 6€ flat fee and an incentive part depending on the decisions made during the experiment. Specifically,

³⁰The experiment presented here was actually part of a bigger experiment. Specifically, subjects also played other types of games whose content depended on the wave (more information about these games are provided in appendix A.1). In any case, subjects played the induced budget experiment first and had no information about the content of the remainder of the experiment, except for the fact that they would participate in other games. In the end, each subject's payoff was computed as the sum of what was earned within each game, but subjects did not receive any feedback about what they earned in each of them before all were played.

³¹In practice such a form has no legal basis : it is akin to a non-binding oath. Yet, some recent experimental studies emphasize that such oaths can actually be efficient commitment devices (see for instance Jacquemet et al., 2013, 2017, 2020).

a third of subjects was randomly chosen to receive the incentive part at the end of each session. The remaining two thirds was only given the flat fee.

Selecting only a subset of subjects for payment is known as the Between-subject Random Incentive System (BRIS). When combined with paying only one decision out of several (the Within-subject Random Incentive System, WRIS), such a payment scheme is known as the Hybrid Random Incentive System (HRIS: Baltussen et al., 2012).³² In our case, the choice of the BRIS/HRIS was guided by two main reasons. First, local politicians appeared reluctant to get paid, which entailed a risk of participation bias. Similarly, the people we reached to contact our target group (e.g., AMF or intermunicipal governments staff) were often embarrassed by this particular feature, which is actually even more crucial for the experiment's success.³³ Second, computing the valid opportunity cost for Politicians proved difficult.³⁴ The risk then being that standard payoffs based on hourly student wage or French median net income appear not salient enough to this specific population. As a consequence, we resort to the BRIS as a way to reduce possible participation bias while providing the opportunity to play with seemingly higher stakes than if every subject was paid.³⁵

Obviously, using the BRIS/HRIS bears the risk of diluting the incentives, hence distorting behavior. In particular, it may increase the propensity to pick a decision at random due to an insufficient payoff saliency (for a review of possible theoretical effects of the payment scheme on experimental measures, see Baltussen et al., 2012). To date, and despite the rising interest in such payment scheme following the surge of massive online experiments, there exists few evidence as regard the empirical

³²Examples of studies using the BRIS or a combination of WRIS and BRIS in a distributional context include Ashraf et al. (2006), Carpenter et al. (2008), Fong and Luttmer (2011), Frigau et al. (2019), Kerschbamer and Müller (2020), among others.

³³For a discussion on the importance of initial contact and efficient networking for a study sollicitating politicians, see the "good practices" review by Vis and Stolwijk (2020).

³⁴In practice, most French local Politicians are volunteers and receive a salary from their private activity, which can be anything. Yet, some may also receive a service allowance which is computed based on standardized national rules or local conventions, and upon which little aggregate data is available. Furthermore, an important part of them are retired (about 45% for Mayors, see Foucault and Bono, 2020).

³⁵Alternatively, we could have implemented a payment scheme based on subjects waiving their payoff to the benefit of a charity, as in Linde and Vis (2017) (also see Butler and Pereira, 2018). However, such a feature could have distorted our control groups' incentives and would consequently have complicated the comparison of our results with the existing literature.

impact of the BRIS/HRIS on observed behavior. Charness et al. (2016) review the existing literature and conclude that the loss of motivation appears actually small in most cases, especially in simple decision environments (Baltussen et al., 2012, reach a similar conclusion in the risk domain). Considering that our experimental task is displayed in a rather intuitive way which entails a direct visual connection between choices and possible payoffs (see the next sub-section), it is likely that the authors' observations holds in the current experiment. Moreover, the limited impact of the payment scheme in a distributional context is also underlined in the recent study by Clot et al. (2018). The authors compare behavior in a dictator game with varying payment scheme and do not find any statistically significant behavioral difference between the pay-all scheme and the BRIS (with either low or high stakes). Only does the hypothetical choice treatment generate a significant change. We consequently expect the payment scheme to have little impact on our results.

The table 1.2 provides an overview of the socio-demographic characteristics of our groups. We find little statistically significant difference between Politicians and Non-Students at the 10% level, except for education level and income: the former is more educated and earns more than the latter (Fisher's exact test: $p = 0.042$ and $p = 0.003$, respectively). The latter difference is however unsurprising considering that both variables are likely correlated and that there exists some evidence that local French politicians are slightly more educated than the average citizen. For instance, in their national survey with Mayors, Foucault and Bono (2020) find that 57% of Mayors continued their studies after the Baccalauréat, that 28% obtained a Bac +2/3 and that 21% obtained at least a Bac +4. According to INSEE statistics, the respective proportions within the general population aged 35-64 are about 50%, about 15% and about 19% (2018). The absence of difference in age, gender and occupation is also expected given that those variables were precisely used during the recruitment process in order for the proportions to match between the two groups. In practice, we had to refuse participation from some female and from some young volunteers. Such an issue was predictable since local politicians in France are about 10 years older than the average citizen and much more likely to be a man. Proportions are 60% male among local politicians (data from the French Ministry of the Interior on the 2014 local elections) against 48% among the general

Table 1.2: Socio-demographic characteristics of the groups

	Pol(Lab)	Pol(Field)	Pol(Total)	NSC	SC
<u>Age</u> [†]	57.91	55.86	56.85	55.21	20.98
<u>Gender</u>					
Male	0.52	0.64	0.58	0.46	0.41
Female	0.48	0.36	0.42	0.53	0.59
<u>Field of Studies</u> [†]					
Fundamental Sciences	0.18	0.46	0.32	0.32	0.13
Econ	0.22	0.26	0.24	0.25	0.28
Humanities	0.13	0.10	0.12	0.15	0.23
Other	0.46	0.18	0.32	0.26	0.37
<u>Education Level</u> ^{*†}					
≤ Bac	0.22	0.46	0.35	0.52	0.45
Bac +3	0.28	0.26	0.27	0.16	0.32
≥ Bac +5	0.37	0.22	0.29	0.26	0.17
Other	0.12	0.06	0.09	0.04	0.07
<u>Occupation</u>					
Private-sector worker	0.25	0.29	0.27	0.22	.
Public-sector worker	0.18	0.24	0.21	0.27	.
Self-employed	0.04	0.10	0.07	0.07	.
Unemployed/Inactive	0.12	0.08	0.10	0.09	.
Retired	0.40	0.29	0.35	0.34	.
<u>Income</u> [*]					
< 2000 €	0.28	0.33	0.31	0.40	.
2000 € ≤ ... ≤ 3000 €	0.40	0.32	0.36	0.43	.
> 3000 €	0.28	0.31	0.29	0.10	.
Do not want to answer	0.03	0.04	0.04	0.07	.
Observations	67	72	139	91	123

Note: Entries are either means (continuous variables, identified by *) or proportions (categorical variables).

E.g., the average Politician is 56.85 years old and there are 58% males among Politicians.

A dot indicates that the variable was not measured for the corresponding group.

Marked variables (†) were used during the recruitment procedure for the Non-student Control group.

French population (from 2014 INSEE statistics).³⁶ We also note that our sample of

³⁶That politicians are generally older, wealthier and more likely to be male than the remainder of the population is not specific to France. As reviewed by Gulzar (2021), age, gender and wealth (as well as ethnicity) are typical markers of the political class in most countries.

politicians actually falls close to the national population of local politicians in terms of age (average 50.52 years old in 2014) and gender composition (60% male).³⁷

Similarly, we observe few differences between Politicians from the Lab and Politicians from the Field. In other words, if there is a selection-into-the-laboratory effect within our group of interest, it does not importantly relate to standard observables or affect both field and lab sessions in the same way. Yet, Politicians from the Field appear less educated and more often have a fundamental science background than Politicians from the Lab (Fisher's exact test: $p = 0.019$ and $p < 0.000$, respectively).³⁸ Although we do not see an obvious explanation for the latter difference, the difference in education level is compatible with a rationale for a selection effect. Indeed, if coming into the lab is costly (both economically and psychologically), we should observe a bigger representation of people facing a lower cost or a greater benefit in doing so in the Lab as compared to the Field. One possibility, which echoes some anecdotal evidence obtained during cheap talks with subjects at the end of sessions, is that the Politicians who came into the Lab had some sort of prior experience with scientific research (like having children currently going through their PhD, which correlates with a personal high degree). Such an experience could in turn facilitate registering to our experiment.

Finally, and predictably, we uncover statistically significant differences between Politicians and Students at least at the 10% level for every available observable. The same is observed between Students and Non-Students, except for gender (Fisher's exact test: $p = 0.486$).

³⁷Nevertheless, we note that our global sample of politicians comprise a higher number of Mayors/intermunicipal councillors than what is observed from the national statistics supplied by the French Ministry of the Interior (respective proportions were 6.99% and 9% in 2014, against 22.79% and 40% in our sample). Indeed, one may argue that local politicians that are particularly motivated by local politics are more likely to be willing to participate to a research project whose purpose is to better understand the very thing that they are passionate about. And these politicians might be more likely to be at the top of local political life or be also invested at other administrative layers.

³⁸We note that part of the difference in Field of Studies can be explained by a change in the possible answers to the Field of Studies question. The end-of-experiment questionnaire in the Field indeed included a wider range of disciplines (hence the smaller "Other" category). However, the difference with the Lab remains statistically significant when we neutralize the difference in possible answers, though it is of lesser importance (Fisher's exact test: $p = 0.033$).

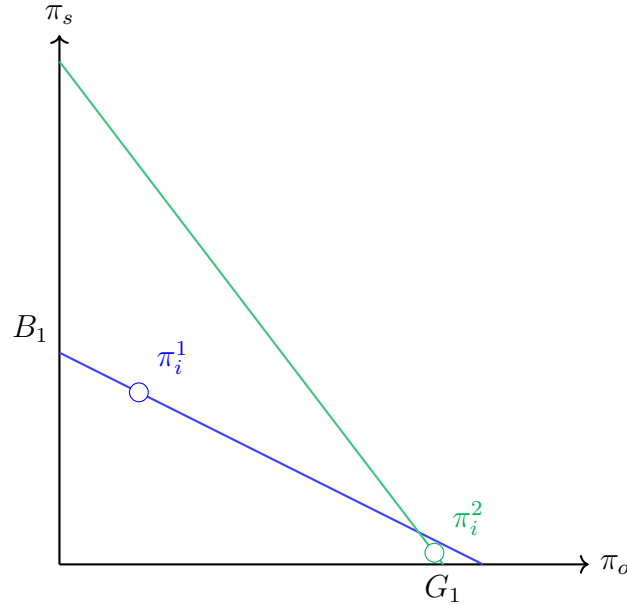
3.2 The experimental task

Our experimental task consists in the induced budget experiment originally proposed by Andreoni and Miller (2002). Specifically, each subject i is exposed to a series of decisions $j = 1, \dots, 30$ in which she is asked to select an ECU allocation $(\pi_{i,s}^j, \pi_{i,o}^j) \geq 0$ between herself (s) and another subject (o) at corresponding prices $(p_{i,s}^j, p_{i,o}^j)$ under the normalized budget constraint $B_i^j = p_{i,s}^j \pi_{i,s}^j + p_{i,o}^j \pi_{i,o}^j = 1$. For ease of understanding, we introduced such a game to subjects in a graphical manner (Fisman et al., 2007, Choi et al., 2007b). Each decision is depicted as a choice from a two-dimensional budget line where each coordinate represents an allocation $(\pi_{i,s}, \pi_{i,o})$ on ECU for self (y -axis) and for another randomly-selected subject in the room (x -axis). Subjects are instructed to move their mouse pointer to the desired coordinate and then click on it.³⁹ A help frame (as well as frames below the x -axis and on left of the y -axis) then displays the payoffs corresponding to the chosen allocation. Subjects are able to revise their choice as much as they want until they are satisfied with it and decide to press the validation button.⁴⁰ For each decision, the budget line is randomly selected from the set of possible budget lines, namely all the budget lines which intersects each axis at 100 ECU maximum and which intersects one axis necessarily above 50 ECU if it intersects the other axis below 50 ECU. As a consequence, each decision problem is independent from the others (within and between subjects). The relative price can be computed ex-post as $p_{i,s/o}^j = p_{i,s}^j / p_{i,o}^j = \bar{\pi}_{i,o}^j / \bar{\pi}_{i,s}^j$, with $\bar{\pi}_{i,s}^j$ and $\bar{\pi}_{i,o}^j$ the j -th budget line endpoints. In other words $p_{i,s/o}^j$ corresponds to the inverse of the slope of the budget line: the steeper it is, the lower $p_{i,s/o}^j$ is. The solid lines B_1 and G_1 in figure 1.2 exemplify possible budget lines. Allocations π_i^1 and π_i^2 provide examples of decisions.

³⁹Contrary to Fisman et al. (2007)'s original experiment on distributional preferences, we restrict the decisions to the budget line instead of allowing for decisions to lie within the budget set. Such a feature aims at simplifying the task, having in mind that our target population (politicians) is both unfamiliar with laboratory procedures and hard to recruit. This is also done by Fisman et al. (2017) in a distributional context or by Choi et al. (2007a, 2014) in an uncertain context. The main consequence is the subsequent impossibility to test for violations of monotonicity. Yet, in Fisman et al. (2007), most subjects did not violate budget balancedness (64 out of 76 subjects). On this issue, also see Müller (2019) who reports only few violations of monotonicity in a similar design.

⁴⁰Considering the difference in screen sizes as well as the relative imprecision of finger-selection as compared to mouse-selection, the graphical interface was slightly adapted for the Field sessions. In particular, subjects could use left/right arrows to move the cursor along the budget line over a gap which they could select in a pre-defined list.

Figure 1.1: Examples of decisions



Note: The subject reveals preferring π_i^1 at prices p_i^1 (budget line B_1) and π_i^2 at prices p_i^2 (budget line G_1).

Resorting to such induced budget experiment has several advantages. First, a continuous choice over a budget line involves more statistical information than a discrete choice within a pre-defined list of allocations. It thus provides a more accurate description of subjects' preferences. Second, we expose each subject to thirty modified dictator games.⁴¹ Obtaining numerous individual observations makes it possible to apply revealed preference tests to individual datasets (as exposed in section A.2). On the contrary, most existing studies focus on aggregate measures of behavior based on pooling individual data and consequently fail to finely account for individual heterogeneity (see e.g., Müller, 2019, on this issue). Third, the randomization of budget lines makes it less likely that a specific decision environment is driving subjects' choices and guarantees that subjects are exposed to many different

⁴¹We reduce the 50-decisions study from Fisman et al. (2007) by two fifths in order to economize on time while preserving an important number of individual decisions. Reducing the number of decisions in a similar design is also implemented by Choi et al. (2014) who replicate the 50-decisions study on risk preferences by Choi et al. (2007a) but with only 25 budget lines per subject.

situations. The latter feature is often missing in real-world situations, as actual budgets for a given agent are not likely to vary importantly over time. In turn, this makes it likely that even random behavior often passes the test (hence a low power, see Beatty and Crawford, 2011, Dean and Martin, 2016). On the contrary, our design provides a powerful test of revealed preference conditions by exposing each subject to many diverse budget constraints (see Choi et al., 2007b). Fourth, an experiment naturally satisfies a falsification requirement entailed by the revealed preference tests. Indeed, observing only a subset of goods does not allow to falsify the utility maximization hypothesis when prices and expenditures on unobserved goods are not fixed (Varian, 1988, Cox, 1997, van Bruggen and Heufer, 2017). Although observational data may provide a large description of public expenditures, it is likely that part of the demand is unobserved. At the same time, especially since studies on observational data are generally panel-structured, it is also likely that unobserved prices and expenditures vary. On the contrary, the number of goods in an experiment is by definition limited but the fixed condition on unobservables is naturally satisfied as the world outside the lab is likely unchanged over the course of the experiment. This makes an experiment a highly valuable tool to test for the consistency with rational behavior (see van Bruggen and Heufer, 2017, for a discussion on this specific point). Fifth, the graphical interface is both intuitive and identical to the standard decision problem in consumer theory. The former feature facilitates subjects' understanding of the game. The latter feature facilitates the comparison of theoretical perspectives and observed behavior, thereby improving the readability of the results.

4 Revealed Preference Analysis

4.1 The Generalized Axiom of Revealed Preferences (GARP)

By exposing each subject to multiple linear budget constraints with varying relative prices, our experiment allows the use of demand analysis's tools to assess whether each individual data is compatible with utility maximization. Specifically, classical revealed preference theory provides a straightforward test based on the seminal theorem by Afriat (1967) (and further refined by Varian, 1982): given a finite collection of budget sets, observed decisions can be rationalized by a well-behaved (piecewise linear, continuous, monotonically increasing and concave) utility func-

tion if and only if they comply with the Generalized Axiom of Revealed Preference (GARP). Formally, write $\pi_i^j R_D \pi_i^k$ if $p_i^j \pi_i^j \geq p_i^k \pi_i^k$ (direct preference relation) and $\pi_i^j P_D \pi_i^k$ if $p_i^j \pi_i^j > p_i^k \pi_i^k$ (strict direct preference relation). GARP then posits: $\pi_i^j R_D \pi_i^k \Rightarrow \neg \pi_i^k P_D \pi_i^j$. That is, GARP states that if we observe allocation π_i^j to be revealed preferred to another allocation π_i^k , then we should not also observe π_i^k to be strictly revealed preferred to π_i^j (i.e., there is no preference cycles).⁴² A subject whose data is consistent with GARP will thus be considered "rational". The figure 1.1 illustrates a textbook GARP violation: subject i chose π_i^1 when π_i^2 was affordable and conversely, so that both allocations are strictly revealed preferred to each other.

4.2 Rationality measures

Yet, GARP provides an exact test of utility maximization (pass or fail) which does not account for the possibility that even globally-consistent subjects are prone to small errors (for instance due to mouse imprecision or fatigue). Furthermore, in the case of a between-subjects study, knowing how far subjects from different groups fall from GARP consistency is a more valuable piece of information than simply assessing how many subjects are GARP-consistent in each group. A continuous measure of rationality in the sense of the goodness-of-fit of the utility maximization model is thus desirable.

The most prominent measure is the Afriat's Efficiency Index (AEI, Afriat, 1972, 1973), which quantifies the fraction by which all budget lines must be shifted in order to remove all GARP violations.⁴³ Formally, let $\tilde{e}_s = (e_i^j)_{j=1}^{30}$ a vector of numbers for subject i with $e_i^j \in [0, 1] \forall j$ the "efficiency level" associated with observation j . Define "relaxed" preference relations $\pi_i^j R_D(\tilde{e}_i) \pi_i^k$ if $e_i^j p_i^j \pi_i^j \geq p_i^k \pi_i^k$ and $\pi_i^j P_D(\tilde{e}_i) \pi_i^k$ if $e_i^j p_i^j \pi_i^j > p_i^k \pi_i^k$. A "relaxed" version of GARP is then $\text{GARP}(\tilde{e}_i)$: $\pi_i^j R(\tilde{e}_i) \pi_i^k \Rightarrow \neg \pi_i^k P_D(\tilde{e}_i) \pi_i^j$. The AEI is finally defined as the largest number e_i^* such that dataset for subject i satisfies $\text{GARP}(\tilde{e}_i^*)$, with \tilde{e}_i^* a vector where $e_i^j = e_i^* \forall j$. Subject i is then said to pass the GARP test at efficiency level \tilde{e}_i^* . The AEI is bounded between zero and one, with indices closer to one indicating a behavior

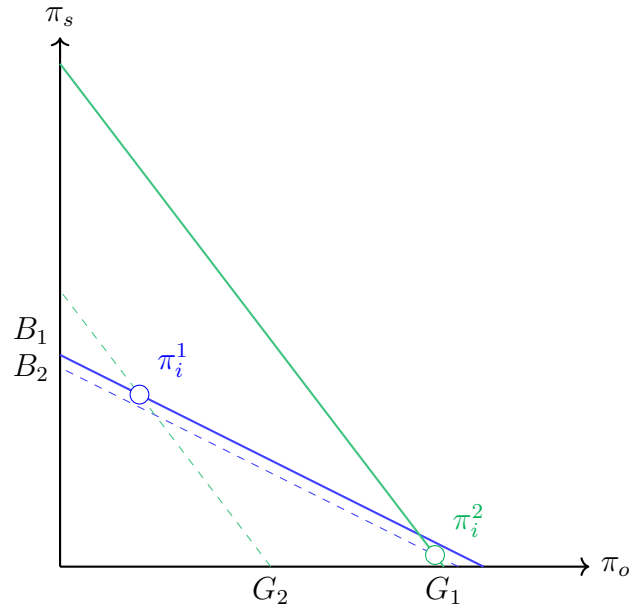
⁴²Such a statement of GARP follows Banerjee and Murphy (2006) who showed that, in a two-commodities space, it is equivalent to Varian (1982)'s original statement which involved the indirect preference relation R (defined as the transitive closure of R_D).

⁴³Such a measure is also commonly known as the Critical Cost Efficiency Index (CCEI).

closer to GARP consistency. The difference between the AEI and one (strict rationality) can thus be interpreted as the fraction of the endowment that the subject is wasting by making inconsistent choices.

The figure 1.2 illustrates the AEI's construction. If the budget line B_1 is shifted through π_s^2 or the budget line G_1 is shifted through π_i^1 , the violation is removed. The latter entails however a much bigger adjustment than the former, so that it appears more convincing to conclude that subject s truly prefers π_i^2 over π_i^1 rather than the converse. The AEI ultimately corresponds to budget adjustment B_2 . In practice, we compute the AEI following the procedure described in Afriat (1987) (also see Murphy and Banerjee, 2015).

Figure 1.2: Graphical example of AEI's construction for a simple GARP violation



Note: The subject reveals preferring both π_i^1 to π_i^2 at prices p_i^1 and π_i^2 to π_i^1 at prices p_i^2 , hence a preference cycle $\pi_i^1 P_D \pi_i^2 P_D \pi_i^1$. Such GARP violation can however be removed by (minimally) shifting budget line B_1 towards B_2 .

Given its empirical popularity, mainly supported by its tight connection to economic theory and its computational simplicity, we rely on the AEI as our primary

rationality measure.⁴⁴ However, the AEI does also present some drawbacks. In particular, the AEI is stringent on sizeable violations but allows multiple violations of GARP as soon as they are sufficiently small. Yet, it seems quite complicated to assess whether few big violations of rationality are worse or better than many small violations. Furthermore, the AEI summarizes violations of different intensity into one single number. As such, it does not provide any information as to which observations generate the biggest violations, and more generally it is only a lower bound on efficiency. In order to account for these drawbacks, we also report two additional indices which are frequently referred to in the literature (although much less empirically resorted to). One methodological feature of the current study indeed is to gauge deviations from rationality based on three different conceptions of "closeness" to the model, whereas the majority of papers mainly deal with (or emphasize) the AEI.⁴⁵

The first additional index is the Houtman and Maks (1985)'s Index (HMI), which is the ratio between the greatest number of observations containing no GARP violation and the total number of observations. The higher the ratio, the less removals will it need to obtain a GARP-consistent dataset, and thus the closer the subject is to maximizing utility. Consequently, the HMI handles big and small violations of GARP identically. Formally, the HMI can be obtained by restricting the efficiency levels e_i^j to be dummy variables (Heufer and Hjertstrand, 2015, Halevy et al., 2018). The HMI is then the maximal fraction of non-zero elements in vector $\tilde{\mathbf{e}}_i$ such that the dataset for subject s complies with $\text{GARP}(\tilde{\mathbf{e}}_i)$. In practice, we compute the HMI by applying Gross and Kaiser (1996)'s algorithm to GARP violations, as suggested by Heufer and Hjertstrand (2015).

The second additional index is the Varian (1990, 1993)'s Efficiency Vector (VEV), which emphasizes the minimal adjustment required to remove all GARP violations associated with each observation. Consequently, it measures the efficiency loss implied by each individual decision and thus identifies where the inefficiencies are the

⁴⁴Studies focusing on the AEI to gauge subjects' rationality include for instance Sippel (1997), Mattei (2000), Andreoni and Miller (2002), Février and Visser (2004), Choi et al. (2007a, 2014), Fisman et al. (2007, 2015a,b, 2017), Müller (2019) and Dzielwski (2020).

⁴⁵There also exists other indices such as the the Money-Pump Index (Echenique et al., 2011), the Swap Index (Apesteguia and Ballester, 2015) or the Minimum Cost Index (Dean and Martin, 2016). However, we stick to the standard practice within the experimental literature on revealed preference for the sake of results' comparability. For a recent review of existing indices and a comparison between them, also see Halevy et al. (2018).

greatest and the smallest. Formally, the VEV corresponds to the vector of efficiency levels e_i^j without the restriction that $e_i^j = e_i^* \forall j$. In other words, budget adjustments are allowed to differ across observations. The VEV provides a broader description of the closeness to utility maximization than the above two indices. Yet, the VEV is somewhat harder to interpret than single numbers such as the AEI or the HMI. In order to account for such difficulty, Varian (1993) proposes taking the minimum in the VEV as a summary of individual rationality and a complementary measure with respect to the AEI. Furthermore, the VEV is computationally more intensive than previous indices, which is the main reason put forward by existing studies to discard it. We compute each subject’s VEV from a mixed-integer linear programming (MILP) procedure.⁴⁶ Indeed, the VEV problem can be expressed as a system of linear expressions, which can be solved using optimization routines (for a detailed presentation of the problem’s formulation, see Heufer and Hjertstrand, 2019). A MILP formulation is convenient because: i) it has an exact solution, ii) any local solution is a global one, and iii) it is a commonly-accepted fix to computationally-intensive problems.

We compute all three indices for each subject and use them to compare how close subjects fall from rational behavior (in the sense of utility maximization) across our experimental groups. For the sake of parcimony, we mainly expose the results based on the AEI, presenting related results for the other two indices in the appendix. Details about each index’s computation are exposed in appendix A.2.

5 Results

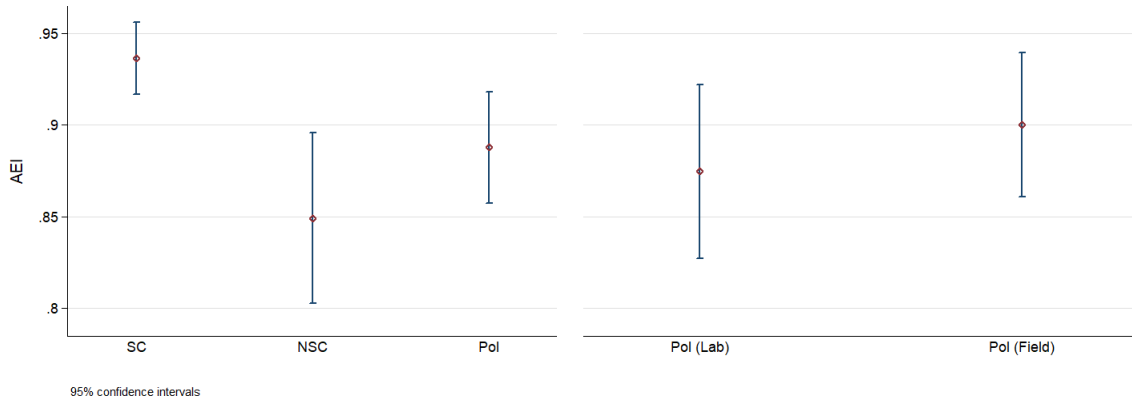
5.1 Between-groups comparisons on rationality

The figure 1.3 plots the mean AEI along with the 95% confidence interval for all the experimental groups. The table 1.3 reports the p-values from several statistical tests across the groups. We report three main statistical tests to assess the sta-

⁴⁶Such class of models was introduced in the revealed preference literature by Cherchye et al. (2008). It provides an alternative to existing iterative algorithms used to compute the VEV (Varian, 1990, 1993, Alcantud et al., 2010, Halevy et al., 2018). Ultimately, these algorithms can only provide a reasonable approximation rather than an exact solution due to the complexity of the problem (Smeulders et al., 2014).

tistical significance of the differences between our groups: the standard parametric t-test (mean differences), the non-parametric robust rank-order test from Fligner and Policello (1981) (median differences)⁴⁷, and the non-parametric Epps-Singleton test (differences in entire distributions).⁴⁸

Figure 1.3: Afriat's Efficiency Index by groups



(a) Main groups

(b) Politicians sub-groups

First, the large majority of Politicians violate GARP at least once (74.82%). However, deviations from strict rationality appear relatively minor since subjects generate an average endowment waste of about 11% (average AEI of 0.888). Besides, removing violations requires only removing between 2 and 3 allocations on average from individual datasets (see the figure A.3 in appendix A.2). In addition, there is no conspicuous difference between Politicians from the Lab (0.875) and

⁴⁷Based on a Monte-Carlo exercise, Feltovich (2003) finds that the Fligner-Policello test outperforms the standard Mann-Whitney test when the sample size is "large" (above 40, which is the case of our study) or when the observed variable is differently dispersed among the compared populations (which is also likely the case). However, we obtain virtually identical results (unreported) when implementing the Mann-Whitney test.

⁴⁸We favour the Epps-Singleton test over the Kolmogorov-Smirnov test when focusing on non-central tendencies because the former is suitable for both continuous and discrete variables (and any mixture of the two) (see Epps and Singleton, 1986). Indeed, the HMI is a discrete variable whereas both the AEI and a summary statistic of the VEV are continuous variables. Comparing experimental groups based on the distributions of all three variables would consequently require a statistical test that can handle both types of variables.

Politicians from the Field (0.9) in the figure 1.3b (that is panel (b) in the figure 1.3), which is confirmed by the statistical testing on mean, median and entire distribution differences with p-values systematically far above conventional levels in the table 1.3. We uncover similar results when scrutinizing the proportion of GARP violations (HMI) or when allowing for budget adjustments to vary across observations for a single subject (VEV). These results are available in appendix A.2 and corresponding p-values are reported in table 1.3. Consequently, no matter how "rationality" is defined, our Politicians sub-groups do not differ significantly. Based on such an observation, we focus in the remainder of our exposure on the "Politicians" group which pools the two sub-groups.

Result 1 (utility maximization from politicians): *Politicians often deviate from rational behavior, but deviations remain limited. Field and Lab Politicians do not differ in terms of closeness to utility maximization.*

Table 1.3: *p*-values from statistical tests for treatment effects on rationality

	Pol v. NSC			Pol v. SC			Lab v. Field		
	AEI	HMI	VEV	AEI	HMI	VEV	AEI	HMI	VEV
t-test	.151	.707	.095	.01	.015	.014	.409	.881	.15
FP test	.363	.883	.428	.017	.028	<0.01	.893	.993	.855
ES test	.65	.955	.609	.034	.043	.132	.547	.582	.31
Observations	230	230	229	262	262	261	139	139	138

Note: AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector. t-test: two-samples Student test. FP: Fligner-Policello. ES: Epps-Singleton. VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993). Similar results (unreported) are observed when comparing the average element in the efficiency vector.

Second, Politicians waste 5 percentage points more money than Students (0.937) and about 4 percentage points less than Non-students (0.849), as emphasized in the figure 1.3a (that is panel (a) in the figure 1.3). However, only the difference with respect to Students is statistically significant. Indeed, Politicians and Non-Students display similar patterns in terms of closeness to utility maximization: they

are respectively 25.18%/27.47% to abide by GARP (against 37.4% for Students), and 66.19%/61.54% to have an index above 0.9 (against 81.3% for Students). Such a result is confirmed by non-parametric tests yielding no evidence of a difference in neither median nor overall distribution between Politicians and Non-Students in table 1.3, contrary to what we observe between Politicians and Students. We uncover similar results when scrutinizing the other two rationality indices (see appendix A.2). Specifically, Politicians and Non-Students require removing on average 2 to 3 allocations, whereas only 1 to 2 such removals are necessary for Students. Furthermore, the average distributions of efficiency levels in the VEV are fairly close between Politicians and Non-Students, and generally lower than the ones for Students.⁴⁹

As an additional robustness check, we also test for differences between Politicians from the Field and both control groups. The results are unchanged regarding the AEI and the HMI: all tests yield non-significant differences between Politicians from the Field and Non-Students whereas all differences are significant in the comparisons involving Students.⁵⁰ The results are less clear-cut regarding the minimum in the VEV. We indeed find a significant mean difference between Politicians from the Field and Non-Students (t-test: $p = 0.037$), whereas the difference between the former and Students is no longer significant (t-test: $p = 0.192$). However, non-parametric tests on medians are consistent with previous observations (FP tests: $p = 0.394$ and $p = 0.029$, respectively).⁵¹ As a consequence, despite some slight discrepancies, our results emphasize the absence of a substantial difference between Politicians and Non-Students. However, we acknowledge that the existing discrepancies all point

⁴⁹We acknowledge that we find a weakly-significant mean difference between Politicians and Non-Students with the VEV. Such a result can be partly attributed to one observation that had to be dropped when computing the VEV: one subject (a Politician from the Field) chose to give nothing to both roles once and was not consistent over his choices. Such a behavior resulted in wasting the full budget and impinging on a strictly positive constraint for the e_i^j in the MILP procedure. Given both non-parametric tests and the tests on the mean element in the VEV do not show significant differences across groups, we do not consider such weakly-significant difference to support enough evidence that Politicians and Non-Students differ with respect to choice consistency in our data (also see the results from the regression analysis below).

⁵⁰Pol v. NSC: AEI (HMI): t-test: $p = 0.11$ (0.695) ; FP test: $p = 0.372$ (0.887) ; KS test: $p = 0.31$ (0.88).

Pol v. SC: AEI (HMI): t-test: $p = 0.069$ (0.044) ; FP test: $p = 0.043$ (0.058) ; KS test: $p = 0.092$ (0.074).

⁵¹Again, these contrasted results may be linked to the inconsistent subject that could not be included in the procedures for computing the VEV.

towards the same direction, so that the overall impression is that Politicians are no less rational than a comparable group (if not slightly more).

Result 2 (utility maximization in Politicians v. Non-Students): *Politicians are at least as consistent in their choices as a comparable control group.*

Result 3 (utility maximization in Politicians v. Students): *Politicians are farther away from utility maximization than a student control group.*

Third, despite between-groups heterogeneity, most of our subjects fall close to utility maximization. Overall, the observed pattern is consistent with existing results, which we interpret as a confirmation that our experiment yields results similar to existing ones. For instance, Fisman et al. (2017) report an average AEI of 0.862 based on an experiment with a representative sample from the American population, which is comparable to our 0.895 average. Focusing on our convenience student sample, we uncover that the average AEI (0.937) is only slightly above the average of 0.899 reported in the original student-based experiment by Fisman et al. (2007). The direction of this difference was indeed predictable since our experiment involved a lower number of budget lines (30 against 50), hence less opportunities to violate GARP. Furthermore, given the 0.8 threshold that prevails within the literature as indicative of rational-enough behavior, we can classify 81.29% Politicians, 73.63% Non-Students and 89.43% Students as almost utility maximizers. Again, such numbers are comparable to what is usually reported in the literature. The same applies to the HMI and the VEV. In particular, the statistics for Students reported in the table A.2 in appendix A.2 based on various elements of the efficiency vector appears quite close to the corresponding table 3 presented in Heufer and Hjertstrand (2019). Besides, the average mean of efficiency levels in the VEV is above 0.99 for all groups. This means that the average observed behavior in each group is 99% as efficient as optimizing behavior, hence further consolidating the compatibility of our dataset with revealed preference axioms.

Result 4 (overall closeness to rational choice): *Overall, the majority of subjects show only minor deviations from utility maximization.*

Such findings are all the more interesting considering the unusual HRIS payment scheme that we implemented: they point towards the little apparent impact of such scheme on choices' consistency. Indeed, had payoffs' saliency decreased dramatically

due to the specific payment scheme, we would have expected much more deviations from GARP than what we actually observe. Although there exists some studies focusing on the impact of the BRIS/HRIS on observed behavior in experimental games (see the literature reviews in Charness et al., 2016, Clot et al., 2018) and despite the fact that the HRIS is increasingly resorted to in survey experiments, we are unaware of any study specifically addressing the impact of the HRIS on rational behavior. Though the present experiment is not designed to directly tackle such an issue, our results still appear encouraging.

5.2 Regression framework

We now turn to an econometric analysis of the differences in rationality measures between our groups, to assess the validity of our results when adding control variables. Indeed, the descriptive statistics from section 3.1 highlight some significant differences in observables between our groups, which may in turn bias our results. Tables 1.4 (main groups) and 1.5 (Politicians sub-groups) present the results of regressing a group variable (dummy) on our rationality indices, along with control variables. Given the nature of our rationality measures and the subsequent concentration of data at 1 (highest rationality score), we consider a right-censored tobit specification. For each rationality index, we display two models for pairwise comparisons between groups. The first model includes only the group variable (taking value 1 for Politicians in table 1.4 and for Politicians from the Field in table 1.5), whereas the second model includes "relevant" control variables. We consider "relevant" the variables for which we observe a statistically significant difference across our groups, namely education level and income (Pol v. NSC) or education level and field of studies (Pol v. SC and Politicians from the Lab v. Politicians from the Field). We also systematically account for gender and age (when available), in view of their ability to explain a wide range of behavior in distribution games.⁵² The first model for each rationality index is exposed in the first column (both tables) as well as in the third column (table 1.4), while the second model is exposed in the second column (both tables) as well as in the fourth column (table 1.4). For each rationality index in

⁵²The number of observations between the models without and with control variables may differ because one subject (a Non-Student) had to leave the laboratory before completing the socio-demographic questionnaire. As a consequence, his characteristics are unobserved.

the table 1.4, the first two columns compare Politicians and Non-Students (NSC), whereas the last two columns compare Politicians and Students (SC).

Table 1.4: Tobit on rationality measures (main groups)

	AEI				HMI				VEV			
	(NSC)	(NSC)	(SC)	(SC)	(NSC)	(NSC)	(SC)	(SC)	(NSC)	(NSC)	(SC)	(SC)
<u>Pol</u> (Baseline: Control)	0.046	0.044	-0.034***	-0.047***	0.003	0.001	-0.017***	-0.021***	0.051	0.054	-0.040***	-0.052***
	(0.033)	(0.035)	(0.012)	(0.013)	(0.012)	(0.013)	(0.005)	(0.006)	(0.037)	(0.037)	(0.013)	(0.014)
<u>Age</u>		-0.002				-0.001				-0.004**		
		(0.002)				(0.001)				(0.002)		
<u>Gender</u> (Baseline: Male)		0.025		-0.020		0.023		-0.019		0.045		0.002
		(0.031)		(0.027)		(0.018)		(0.014)		(0.033)		(0.030)
<u>Higher education</u> (Baseline: No)		0.030		0.044*		0.010		0.019		0.033		0.062**
		(0.036)		(0.026)		(0.016)		(0.014)		(0.043)		(0.029)
<u>Income</u> (Baseline: Between 2000 € and 3000 €)												
< 2000 €		-0.110**				-0.051***				-0.138***		
		(0.050)				(0.018)				(0.049)		
> 3000 €		-0.061				-0.017				-0.078		
		(0.043)				(0.019)				(0.048)		
Do not want to answer		-0.184**				-0.065***				-0.211**		
		(0.085)				(0.024)				(0.085)		
<u>Field of Studies</u> (Baseline: Fundamental Science)												
Economics				-0.088**				-0.012				-0.077*
				(0.037)				(0.019)				(0.039)
Humanities				-0.138**				-0.031				-0.151**
				(0.057)				(0.027)				(0.066)
Other				-0.050*				-0.009				-0.039
				(0.029)				(0.020)				(0.030)
Observations	230	229	262	262	230	229	262	262	229	228	261	261
LR chi2	1.869	10.946	8.320	21.072	0.065	23.061	11.375	17.313	1.916	25.685	10.348	19.537
Prob > χ^2	0.172	0.141	0.004	0.002	0.799	0.002	0.001	0.008	0.166	0.001	0.001	0.003
Bootstrapped p-value	0.192	0.224	0.015	0.002	0.804	0.932	0.003	0.004	0.173	0.142	0.013	0.001

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses.

Note: AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector.

VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993). Similar results (unreported) are observed when regressing the average element in the efficiency vector.

The "Bootstrapped p-value" report the p-value associated with the group variable when applying the score cluster bootstrap correction for the small number of clusters (Kline and Santos, 2012)⁵³. The bootstrap is conducted with 4999 repetitions.

⁵³The score cluster bootstrap procedure is the counterpart of the wild cluster bootstrap procedure (Cameron et al., 2008, Cameron and Miller, 2015) when maximum likelihood estimators are used instead of standard ordinary least squares estimators. It aims at correcting the potential bias on standard errors (hence statistical tests) when there are too few clusters. Indeed, in such a situation, the residuals are downward biased (the model overfits the data) and the confidence intervals are too narrow around the estimated coefficient (which increases the probability of rejecting the null hypothesis).

Table 1.5: Tobit on rationality measures (Politicians sub-groups)

	AEI		HMI		VEV	
<u>Experiment</u> (Baseline: Lab)	0.009 (0.018)	0.001 (0.021)	-0.002 (0.007)	-0.004 (0.009)	0.019 (0.017)	0.019 (0.021)
<u>Age</u>		-0.002 (0.001)		-0.002 (0.001)		-0.002 (0.002)
<u>Gender</u> (Baseline: Male)		0.018 (0.031)		-0.005 (0.022)		0.054 (0.035)
<u>Field of Studies</u> (Baseline: Fundamental Science)						
Econ		-0.170*** (0.054)		-0.035 (0.024)		-0.149*** (0.052)
Humanities		-0.249*** (0.064)		-0.065** (0.025)		-0.253*** (0.075)
Other		-0.060 (0.049)		-0.012 (0.030)		-0.024 (0.045)
<u>Higher education</u> (Baseline: No)		0.037 (0.040)		0.031* (0.018)		0.071* (0.039)
Observations	139	139	139	139	138	138
LR chi2	0.273	36.003	0.068	34.227	1.145	30.938
Prob > χ^2	0.601	0.000	0.794	0.000	0.285	0.000
Bootstrapped p-value	0.621	0.956	0.792	0.693	0.311	0.385

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses.

Note: AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector.

VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993). Similar results (unreported) are observed when regressing the average element in the efficiency vector.

The "Bootstrapped p-value" report the p-value associated with the group variable when applying the score cluster bootstrap correction for the small number of clusters (Kline and Santos, 2012). The bootstrap is conducted with 4999 repetitions.

We observe the same results as before, no matter the rationality index under scrutiny. The sign and magnitude of the dummy variable's coefficient as well as the corresponding p-values are indeed close across models for a given comparison (Pol v. NSC, Pol v. SC or Lab v. Field). First, Politicians are not significantly closer to GARP consistency than Non-Students (table 1.4). Indeed, though slightly above 0, the coefficient on the group dummy is never statistically different from 0 in the "NSC" columns, with or without controls.⁵⁴ Second, a similar conclusion can be drawn from comparing Politicians from the Lab with Politicians from the Field (table

⁵⁴The results do not change upon removing the income variable's "Do not want to answer" category (11 subjects).

1.5): the two groups do not significantly differ.⁵⁵ Third, Politicians are significantly farther away from GARP consistency than Students (table 1.4). Indeed, no matter the specification or the index under scrutiny, we find overwhelming evidence ($p < 0.01$) that Politicians are less rational than Students in our experiment. Finally, the introduction of control variables almost have no effect on the group comparisons.

Regarding control variables, we uncover results consistent with existing findings (although the effects are not always significant): young, highly-educated and wealthy individuals are more likely to behave consistently with utility maximization than other groups. The effect of age is generally small, but it may be explained by the small variations in age that our non-student sample comprises (from 38 to 75 years old). Furthermore, we cannot control for age in the regressions concerning Students due to a strict collinearity issue. We are thus unable to disentangle the effect of the group from the effect of age, and it is likely that Students appear more GARP-consistent than both other groups because they are younger, not because they are students.⁵⁶ Interestingly, the effect of gender is often mentioned in the literature (males being more rational than females), but in our sample the signs of the estimates are not consistent across models or indices (and never significant). On the contrary, we find a strong and positive effect of majoring in fundamental sciences on rational behavior as opposed to other majors.

⁵⁵In the appendix B.5, we test whether additional variables are good predictors of rationality among Politicians. However, we do not find much significant relationships. Such a result is interesting considering that our sample of Politicians slightly differ in terms of politician-specific characteristics from the national sample of politicians (especially the election status and holding an intermunicipal term).

⁵⁶In the appendix B.5, we also explore whether past experience of economic experiments is a predictor of rationality among Students. Indeed, it is also generally an important difference between student samples and non-student samples, as students are more regularly offered the opportunity to participate in economic experiments. Interestingly, we do observe a very significant positive effect of having participated at least twice in the past on all our rationality indices. When excluding such subjects from the Student control group and comparing with Politicians, we find that the group dummy still has a negative effect, but also that the statistical significance is much reduced. In other words, the age effect mentioned in the literature may partly be confounded with an experience effect.

6 Power analyses

6.1 Power of the revealed preference tests

One issue with the revealed preference analysis exposed so far is the one relative to the tests' power. Basically, the rationality indices indicate the adjustments on the data that would have been necessary so that we cannot reject that subjects are maximizing some well-behaved utility function. At the same time, the experimental design might just be too weak to actually detect violations of rationality for given adjustments, even when no adjustments are needed. Consider for instance the case where budget lines never intersect: no matter the observed behavior he would not violate GARP. In other words, the experiment may lack power so that we cannot know whether subjects abide by (relaxed) GARP due to sufficiently-consistent behavior or to power deficiencies.

A related issue stems from the fact that there is no scientific reason *a priori* to prefer one efficiency level over another in order to classify subjects as close enough to rationality and proceed with modelling their behavior within an utility maximization framework. A power analysis precisely provides a way of assessing how much one can afford to relax the revealed preference conditions without hinging too much on power, hence the optimal rationality score.

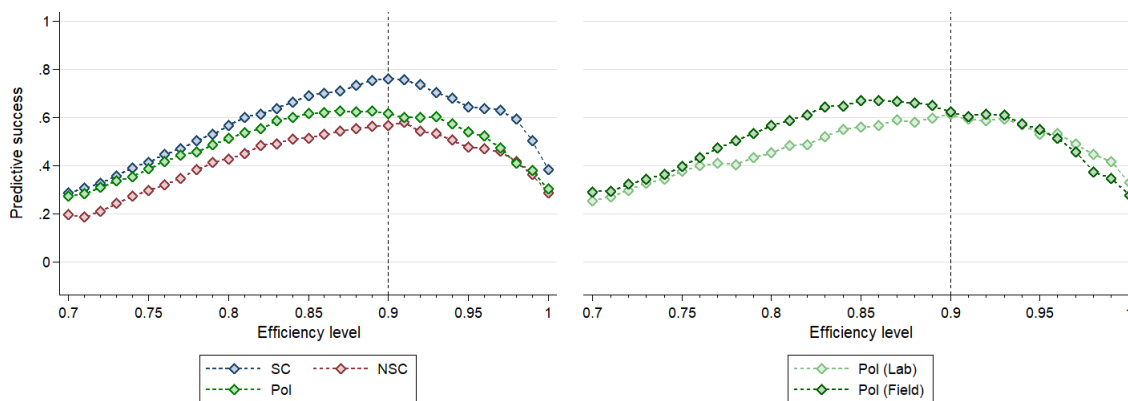
To assess the power of our experimental design in detecting violations of rationality, we rely on an individual measure combining efficiency and power advocated by Beatty and Crawford (2011). Specifically, such a measure is computed as the difference between the pass rate (whether or not the subject pass the $\text{GARP}(\tilde{e}_i^*)$ test) and the relative size of the target area (the ratio between the set of choices that are compatible with revealed preference conditions and the set of all possible choices). The relative size of the target area is inversely related to the power of the design (the tighter the conditions, the higher the power), which can be interpreted as the frequency at which uniformly-random subjects violate GARP when choosing on actual budget sets (see Dean and Martin, 2016, Heufer and Hjertstrand, 2019).⁵⁷ The result is bounded in $[-1, 1]$, with -1 indicating failure to abide by even the loosest

⁵⁷The uniform benchmark may not be the best counterfactual but it naturally applies to any choice environment and is heavily relied on within the literature on revealed preference. See Bronars (1987) for a discussion.

conditions, 1 indicating passing the most stringent test, and 0 indicating the impossibility to discriminate actual behavior from random behavior. Ultimately, such a measure follows the concept of predictive success originally exposed by Selten (1991): the higher the measure, the better the success of the utility maximization model in accurately predicting observed behavior.

We compute the predictive power measure for subject i over several efficiency levels \tilde{e}_i^* as follows: i) check whether or not subject i complies with $\text{GARP}(\tilde{e}_i^*)$ (binary pass rate), ii) simulate a hundred uniform subjects choosing at random on the same thirty budget lines as subject i , iii) compute the proportion of simulated subjects that do not comply with $\text{GARP}(\tilde{e}_i^*)$ (power). Repeating such procedure for various efficiency levels given an arbitrarily fine grid provides a way to identify the efficiency level which maximizes the predictive power (hence a good candidate for defining "rational-enough" behavior). The figure 1.4 summarizes the results on predictive power: for each efficiency level, the graph plots the average predictive power within the corresponding group⁵⁸. We also compute the Selten score for our different rationality indices, defined as the difference between subject i index and the average index in the corresponding pool of simulated subjects (see Dean and Martin, 2016), and report the results in table 1.6.

Figure 1.4: Mean predictive success for different efficiency levels across groups



(a) Main groups

(b) Politicians sub-groups

⁵⁸Similar results are observed based on median predictive power (see figure A.10 in appendix A.4).

Overall, this analysis confirms that the utility maximization hypothesis fits the data quite well on average. Indeed, it outperforms by far an alternative random behavior hypothesis with a maximal average predictive success of 0.66 reached for an efficiency level of 0.9 (figure 1.4).⁵⁹ Predictably, we also recover previous results: the utility maximization model fits Students' data the best (maximal average predictive success of about 0.8) and fits Politicians' data better than Non-Students' data. Interestingly, the maximal average predictive success achieved for Politicians and Non-Students is close (about 0.6 and reached at an efficiency level of about 0.9), which means that there is no conspicuous difference between both groups regarding the optimal efficiency-power tradeoff. Nevertheless, the model's fit is better for Politicians than for Non-Students at virtually all efficiency levels. This is unsurprising considering that Politicians have a greater AEI than Non-Students on average (figure 1.3), but we also note that the predictive success's pattern for Politicians is close to Students' one for low efficiency levels (until about 0.8) and then draws closer to Non-Students' one. One interpretation could then be that Politicians are actually exposed to a slightly more stringent test of revealed preference conditions, as compared to both our control groups. Finally, the predictive power of the model is close between the two groups of Politicians (figure 1.4b). Specifically, the model's fit is better for Politicians from the Field, except for high efficiency levels (0.95 and above), which is consistent with the lower AEI that is observed for Politicians from the Lab (figure 1.3b). Interestingly, the picture based on the median predictive power for each efficiency level (see figure A.10 in appendix A.4) provides a virtually identical pattern for the two groups.

⁵⁹The 0.9 threshold on the AEI to define rational behavior that prevail in our study differs from the 0.8 threshold documented in the literature. The difference is likely due to the reduction in the number of choices that we implement (30 instead of 50), which increases the likelihood to pass the GARP test by chance. In the original study by Fisman et al. (2007), the authors document the 0.8 choice based on Bronars (1987)'s approach to statistical power: they compare the actual distribution of the rationality index with the distribution obtained from simulating 25000 uniformly-random subjects. As an additional robustness check, we also implement such procedure for our indices. In addition, we plot distributions of the indices when bootstrapping on actual choices to construct the simulated data instead of picking one possible allocation at random, following Andreoni and Miller (2002). Such analyses confirm that most of actual subjects are much closer to utility maximization than the vast majority of simulated subjects. For instance, 70.25% of actual subjects abide by GARP(0.9) against only 4.92% of uniformly-random subjects and 4.15% of bootstrap-random subjects (see appendix A.4). For an insightful discussion and a presentation of several ways to address the power issue of an experimental testing revealed preference conditions, see Andreoni et al. (2013).

Table 1.6: Summary statistics for Selten scores on each rationality index across groups

	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
AEI	0.150 (0.198)	0.173 (0.177)	0.162 (0.187)	0.120 (0.226)	0.203 (0.111)	0.165 (0.179)
HMI	0.157 (0.084)	0.157 (0.086)	0.157 (0.085)	0.151 (0.089)	0.174 (0.064)	0.161 (0.080)
VEV (min)	0.176 (0.228)	0.227 (0.182)	0.203 (0.206)	0.140 (0.263)	0.251 (0.128)	0.203 (0.205)
VEV (mean)	0.026 (0.041)	0.037 (0.060)	0.032 (0.052)	0.019 (0.022)	0.037 (0.057)	0.030 (0.049)

Note: The values for each group and each rationality index correspond to the mean and standard deviation (in parentheses) of the Selten score for the corresponding rationality index, computed as the difference between the actual index and the mean index of a hundred simulated uniform-random subjects.

Similar conclusions are derived when looking at the Selten scores for our different rationality measures (table 1.6). In particular, our subjects fall on average closer to adopting rational behavior than uniform-random choosers by approximately 16 percentage points based on the AEI/HMI, and up to 20 percentage points based on the minimum in the VEV (see the Total column). Again, we observe some differences across our groups with a greater difference between simulated subjects and Students than between simulated subjects and other groups, but the Selten scores are always highly significantly different from 0 (t-test: $p < 0.01$ for any group no matter the rationality index under scrutiny). Globally, we are thus confident that our experiment exposes subjects to sufficiently powerful revealed preference tests. Furthermore, the magnitude of the difference in Selten scores between our groups is globally the same as the magnitude of the difference in the corresponding rationality indices, which indicates that our groups are actually exposed to similarly powerful revealed preference tests. Yet, we also observe that the difference in Selten scores are bigger (resp. smaller) than the difference in the corresponding rationality indices for Politicians than for Non-Students (resp. Students). Such a feature

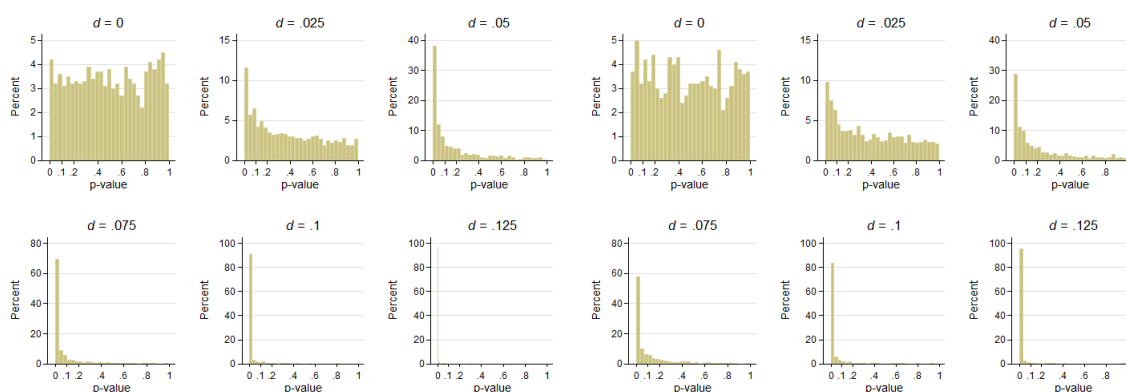
is consistent with Politicians being exposed to slightly tighter revealed preference conditions than both control groups, hence with the idea that Politicians appear slightly more rational than a comparable group. Finally, the differences between the two Politicians sub-groups are consistent with previous results on the rationality indices' levels: Politicians from the Field score higher than Politicians from the Lab, but the differences remain limited (and match exactly the differences in rationality indices).

6.2 Power of the statistical tests

Like our revealed preference tests, that statistical tests supporting our main results on group comparisons are confronted to a power issue: most of the time we fail to reject the null hypothesis of no difference between Politicians and Non-Students as well as between the two Politicians sub-groups. Obviously, failing to reject an hypothesis is not the same as accepting it. In order to shed more light on the power of the statistical tests, we conduct a Monte-Carlo exercise to determine how large the difference d between the groups should be in order for the tests to detect it (the procedure is similar to Müller, 2019, and is detailed in appendix A.4). Specifically, we plot the distribution of a hundred p-values for each pairwise difference and each statistical test in order to determine the magnitude of the difference that is necessary for much of the tests to rule out the null hypothesis of no difference at conventional significance level. The lower the magnitude, the higher the test's power.

The figure 1.5 presents an example for the Fligner-Policello test on median differences in the AEI. Distributions of p-values from the other statistical tests yield similar results (see table A.7 in appendix A.4 for summary statistics). Comparing Politicians and Non-Students, we observe that the overwhelming majority of p-values are below 0.05 for $d = 0.1$ and below 0.1 for $d = 0.075$. In other words, a 7.5 percentage points difference between the groups can be ruled out in our experiment. Actually, smaller differences around 5 percentage points also have a decent chance of being detected (although it is lower than 80%). The p-value distributions for the tests on the differences in the remaining rationality indices are exposed in appendix A.4. Globally the same analysis holds for the minimum in the VEV. Regarding the

Figure 1.5: P-values from simulations of a hundred Fligner-Policello tests on differences in AEI across groups



(a) Pol v. NSC

(b) Lab v. Field

HMI, the tests are actually more powerful: virtually all p-values are below 0.05 at $d = 0.05$ and even a 2.5 percentage points difference is likely detected. The overall pattern is similar when comparing the two Politicians sub-groups, though the tests are slightly less powerful for all tests and all rationality indices. We conclude from such simulation exercise that our tests seem actually quite powerful and that the differences we observe between our groups are likely limited.

7 Concluding remarks

The rational choice framework is the cornerstone of existing economic analyses of political decisions, and also constitutes an important theoretical bedrock in political science as well as in public administration. Such framework derives empirical predictions from assuming that political agents's behavior can be modelled as utility maximizers. The current study offers direct evidence on the predictive success of the utility maximization model by applying standard revealed preference tests to decisions made by actual politicians in a controlled environment. Specifically, we non-parametrically evaluate the consistency of individual-level data with the Generalized Axiom of Revealed Preference (GARP) and quantify the deviations from it using several continuous rationality measures. We find that most politicians violate GARP but only marginally. Only little adjustments on the data are indeed

required to remove the violations. In particular, politicians are not less rational than a comparable control group, and actually waste slightly less money by making inconsistent choices. Furthermore, we obtain similar results whether we run the experiment in the lab or in the field. Results obtained from laboratory experiments on the decision-making ability of politicians are thus likely to provide a fairly consistent description of the overall population of politicians. As a consequence, the utility maximization hypothesis appears a reasonably-useful approximation to observed decisions made by politicians. Additionally, we find a convenience student sample to generate less inefficiencies than politicians. The generalizability of results obtained from studies using a convenience sample to focus on a policy-making issue should thus be interpreted with caution.

That actual policy-makers are as (ir)rational as the remainder of the population obviously has important implications for our understanding of political decisions. Our results point towards the overall usefulness of the rational choice framework in explaining decisions made by politicians. By extension, they also enhance the average efficiency of incentive schemes or institutions that have been proposed to correct politicians' behavior based on the assumption that they act rationally (see e.g., Laffont, 2000). But such results also have major consequences on the the way we, as scientists, communicate on our results to a political audience and make policy recommendations. As emphasized by Berggren (2012), very few papers in behavioral economics actually take into account their very own scientific premisses when whispering into the ears of the Prince. On the contrary, they generally implicitly assume that their interlocutor will behave rationally when both assimilating the scientific information and then implementing the proposed policy. Such a practice may eventually be detrimental to a proper design of policies inspired by scientific knowledge (also see Thomas, 2019). Though comforting in this view, our results also indicate that one cannot expect all politicians to strictly behave rationally, which ultimately means that such communication issue matters. In addition, considering the amounts at stake in the decisions politicians make, even small deviations from rationality may ultimately generate a substantial absolute amount of public funds waste. A careful attention to behavioral insights applied to policy-makers could consequently help design useful incentive schemes or institutions to prevent inefficient decisions.

Obviously, our design does not allow to tackle all interesting issues related to politicians' decision-making ability, which could serve as avenue for future studies.

First, the choice anomalies which politicians have been showed to be subject to in the literature (e.g., Linde and Vis, 2017, Sheffer et al., 2018, Vis, 2019, Sheffer and Loewen, 2019, Baekgaard et al., 2019a) are possible determinants of deviations from GARP. For instance, the adoption of a specific heuristics may cause subject to fail the GARP test. However, such heuristics may also imply that deviations from GARP happen in a limited range so that departures from rationality remain small because they entail some form of regularity in observed behavior. Considering that politicians may be often confronted to using heuristics (for instance due to the complexity of their usual decision environment), it is likely that they more often deviate from GARP but generate smaller violations than a comparable control group whose deviations could be driven by other causes (cognitive biases, fatigue, lack of focus, etc.). The methodology in the current chapter does not allow to accurately disentangle between possible sources of irrational behavior, but it would be interesting to explore the reasons why politicians deviate from GARP and whether such reasons are group-specific.

Second, we do not look into the politicians' representativeness issue, that is the congruence between voters' characteristics and their representatives' ones. Instead, we compare politicians and a group with similar characteristics except the election status and taken from the general population. That politicians are not representative of their constituents may indeed be an issue in democratic regimes (Mansbridge, 2003). At the same time, that the political selection process leads to the election of efficient political elites is certainly a valuable feature that many voters could agree upon. Digging into this kind of questions would indeed require comparing rational behavior from a set of representatives and a set of the corresponding represented population (which at the national level would for instance mean a sample from national politicians and a representative sample from the national population).

Third, our experiment sets aside the fact that politicians are representatives, hence supposed to implement the preferences of the people who they represent (citizens) and not their own preferences. However, implementing the preferences of others and not one's own preferences may not generate the same choice inconsistencies (even when both types of preferences are aligned). At the same time, it may be that politicians are more rational than the remainder of the population when taking into account this dimension of political decisions, because they are used to

decide for others.⁶⁰ Investigating such a feature would thus complement the present work by marginally modifying the decision environment towards the usual political environment.

Fourth, running such kind of experiment with other types of politicians (say national politicians) or targetting local politicians deciding within bigger local structures (big cities, counties) could also provide interesting insights on the way such structures or political careers affect the decision-making ability of policy-makers. Indeed, such analysis could explain why the existing results in the literature are somewhat contradictory: in particular we do not find local politicians to behave significantly more rationally than a comparable group, whereas studies on political elite decision-making often points towards their higher decision-making abilities. Maybe the difference lies on the fact that the political elite under scrutiny is different across studies: local politicians are indeed very close to a comparable group by nature, whereas higher-level politicians differ in many more ways. To date, we are unaware of studies systematically comparing elected representatives' allocation decisions at different political levels.

⁶⁰However, it is debatable whether politicians systematically ponder what citizens would want them to do when making a decision.

References

- Afriat, S. N. (1967). The construction of utility functions from expenditure data. International Economic Review, 8(1):67–77.
- Afriat, S. N. (1972). Efficiency estimation of production functions. International Economic Review, pages 568–598.
- Afriat, S. N. (1973). On a system of inequalities in demand analysis: an extension of the classical method. International Economic Review, pages 460–472.
- Afriat, S. N. (1987). Logic of choice and economic theory. Oxford: Clarendon Press.
- Alcantud, J. C. R., Matos, D. L., and Palmero, C. R. (2010). Goodness-of-fit in optimizing a consumer model. Mathematical and Computer Modelling, 52(7-8):1088–1094.
- Amador-Hidalgo, L., Brañas-Garza, P., Espín, A. M., García-Muñoz, T., and Hernández-Román, A. (2021). Cognitive abilities and risk-taking: Errors, not preferences. European Economic Review, 134:103694.
- Andreoni, J., Gillen, B. J., and Harbaugh, W. T. (2013). The power of revealed preference tests: Ex-post evaluation of experimental design. Unpublished manuscript.
- Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. Econometrica, 70(2):737–753.
- Apestequia, J. and Ballester, M. A. (2015). A measure of rationality and welfare. Journal of Political Economy, 123(6):1278–1310.
- Arceneaux, K., Dunaway, J., and Soroka, S. (2018). Elites are people, too: The effects of threat sensitivity on policymakers’ spending priorities. PloS one, 13(4):e0193781.
- Ashraf, N., Bohnet, I., and Piankov, N. (2006). Decomposing trust and trustworthiness. Experimental Economics, 9(3):193–208.
- Ashworth, S. (2012). Electoral accountability: Recent theoretical and empirical work. Annual Review of Political Science, 15:183–201.

- Axelrod, R. (2015). Structure of decision: The cognitive maps of political elites. Princeton university press.
- Baekgaard, M., Belle, N., Serritzlew, S., Sicilia, M., and Steccolini, I. (2019a). Performance information in politics: How framing, format, and rhetoric matter to politicians' preferences. Journal of Behavioral Public Administration, 2(2).
- Baekgaard, M., Christensen, J., Dahlmann, C. M., Mathiasen, A., and Petersen, N. B. G. (2019b). The role of evidence in politics: Motivated reasoning and persuasion among politicians. British Journal of Political Science, 49(3):1117–1140.
- Baltussen, G., Post, G. T., Van Den Assem, M. J., and Wakker, P. P. (2012). Random incentive systems in a dynamic choice experiment. Experimental Economics, 15(3):418–443.
- Banerjee, S. and Murphy, J. H. (2006). A simplified test for preference rationality of two-commodity choice. Experimental Economics, 9(1):67–75.
- Banerjee, S. and Murphy, J. H. (2011). Do rational demand functions differ from irrational ones? evidence from an induced budget experiment. Applied Economics, 43(26):3863–3882.
- Baudry, M., Leprince, M., and Moreau, C. (2002). Préférences révélées, bien public local et électeur médian: tests sur données françaises. Economie prevision, 156(5):125–146.
- Beatty, T. K. and Crawford, I. A. (2011). How demanding is the revealed preference approach to demand? American Economic Review, 101(6):2782–95.
- Berggren, N. (2012). Time for behavioral political economy? an analysis of articles in behavioral economics. The Review of Austrian Economics, 25(3):199–221.
- Bernheim, B. D. and Rangel, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. American Economic Review, 97(2):464–470.
- Besley, T. (2005). Political selection. Journal of Economic Perspectives, 19(3):43–60.
- Besley, T. and Coate, S. (1997). An economic model of representative democracy. The Quarterly Journal of Economics, 112(1):85–114.

- Besley, T., Folke, O., Persson, T., and Rickne, J. (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. American Economic Review, 107(8):2204–42.
- Besley, T. and Reynal-Querol, M. (2011). Do democracies select more educated leaders? American Political Science Review, pages 552–566.
- Bischoff, I. and Siemers, L.-H. (2013). Biased beliefs and retrospective voting: why democracies choose mediocre policies. Public Choice, 156(1):163–180.
- Borge, L.-E. and Tovmo, P. (2009). Myopic or constrained by balanced-budget rules? the intertemporal spending behavior of norwegian local governments. FinanzArchiv/Public Finance Analysis, pages 200–219.
- Bronars, S. G. (1987). The power of nonparametric tests of preference maximization. Econometrica: Journal of the Econometric Society, pages 693–698.
- Butler, D. M. and Pereira, M. M. (2018). Are donations to charity an effective incentive for public officials? Journal of Experimental Political Science, 5(1):68–70.
- Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics, 90(3):414–427.
- Cameron, A. C. and Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. Journal of Human Resources, 50(2):317–372.
- Caprara, G., Francescato, D., Mebane, M., Sorace, R., and Vecchione, M. (2010). Personality foundations of ideological divide: A comparison of women members of parliament and women voters in italy. Political Psychology, 31(5):739–762.
- Carpenter, J., Connolly, C., and Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. Experimental Economics, 11(3):282–298.
- Caselli, F. and Morelli, M. (2004). Bad politicians. Journal of Public Economics, 88(3-4):759–782.
- Chang, C. and Turnbull, G. K. (2002). Bureaucratic behavior in the local public sector: A revealed preference approach. Public Choice, 113(1-2):191–210.

- Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. Journal of Economic Behavior & Organization, 131:141–150.
- Chaudhuri, A., Iversen, V., Jensenius, F. R., and Maitra, P. (2020). Selecting the best of us? politician quality in village councils in west bengal, india. Technical report, CESifo Working Paper.
- Chen, C.-C., Chiu, I.-M., Smith, J., and Yamada, T. (2013). Too smart to be selfish? measures of cognitive ability, social preferences, and consistency. Journal of Economic Behavior & Organization, 90:112–122.
- Cherchye, L., De Rock, B., Sabbe, J., and Vermeulen, F. (2008). Nonparametric tests of collectively rational consumption behavior: an integer programming procedure. Journal of Econometrics, 147(2):258–265.
- Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007a). Consistency and heterogeneity of individual behavior under uncertainty. American Economic Review, 97(5):1921–1938.
- Choi, S., Fisman, R., Gale, D. M., and Kariv, S. (2007b). Revealing preferences graphically: an old method gets a new tool kit. American Economic Review, 97(2):153–158.
- Choi, S., Kariv, S., Müller, W., and Silverman, D. (2014). Who is (more) rational? American Economic Review, 104(6):1518–50.
- Christensen, J., Dahlmann, C. M., Mathiasen, A. H., Moynihan, D. P., and Petersen, N. B. G. (2018). How do elected officials evaluate performance? goal preferences, governance preferences, and the process of goal reprioritization. Journal of Public Administration Research and Theory, 28(2):197–211.
- Christensen, J. and Moynihan, D. P. (2020). Motivated reasoning and policy information: Politicians are more resistant to debiasing interventions than the general public. Behavioural Public Policy, pages 1–22.
- Cleave, B. L., Nikiforakis, N., and Slonim, R. (2013). Is there selection bias in laboratory experiments? the case of social and risk preferences. Experimental Economics, 16(3):372–382.

- Clot, S., Grolleau, G., and Ibanez, L. (2018). Shall we pay all? an experimental test of random incentivized systems. Journal of Behavioral and Experimental Economics, 73:93–98.
- Cooper, J. C. and Kovacic, W. E. (2012). Behavioral economics: implications for regulatory behavior. Journal of Regulatory Economics, 41(1):41–58.
- Cox, J. C. (1997). On testing the utility hypothesis. The Economic Journal, 107(443):1054–1078.
- Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H., and Zhukova, V. (2016). Cognitive (ir) reflection: New experimental evidence. Journal of Behavioral and Experimental Economics, 64:81–93.
- Dal Bó, E., Dal Bó, P., and Di Tella, R. (2006). "plata o plomo?": Bribe and punishment in a theory of political influence. American Political Science Review, pages 41–53.
- Dal Bó, E., Dal Bó, P., and Eyster, E. (2018). The demand for bad policy when voters underappreciate equilibrium effects. The Review of Economic Studies, 85(2):964–998.
- Dal Bó, E. and Di Tella, R. (2003). Capture by threat. Journal of Political Economy, 111(5):1123–1154.
- Dal Bó, E. and Finan, F. (2018). Progress and perspectives in the study of political selection. Annual Review of Economics, 10:541–575.
- Dal Bó, E., Finan, F., Folke, O., Persson, T., and Rickne, J. (2017). Who becomes a politician? The Quarterly Journal of Economics, 132(4):1877–1914.
- Dean, M. and Martin, D. (2016). Measuring rationality with the minimum cost of revealed preference violations. Review of Economics and Statistics, 98(3):524–534.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. Journal of Economic Literature, 47(2):315–72.
- Dietrich, B. J., Lasley, S., Mondak, J. J., Remmel, M. L., and Turner, J. (2012). Personality and legislative politics: The big five trait dimensions among us state legislators. Political Psychology, 33(2):195–210.

- Downs, A. (1957). *An economic theory of democracy*. New York: Harper & Row.
- Dziewulski, P. (2020). Just-noticeable difference as a behavioural foundation of the critical cost-efficiency index. *Journal of Economic Theory*, page 105071.
- Echenique, F., Lee, S., and Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223.
- Epps, T. and Singleton, K. J. (1986). An omnibus test for the two-sample problem using the empirical characteristic function. *Journal of Statistical Computation and Simulation*, 26(3-4):177–203.
- Esaiasson, P. and Öhberg, P. (2020). The moment you decide, you divide: How politicians assess procedural fairness. *European Journal of Political Research*, 59(3):714–730.
- Fatas, E., Neugebauer, T., and Tamborero, P. (2007). How politicians make decisions: A political choice experiment. *Journal of Economics*, 92(2):167–196.
- Feltovich, N. (2003). Nonparametric tests of differences in medians: comparison of the wilcoxon–mann–whitney and robust rank-order tests. *Experimental Economics*, 6(3):273–297.
- Février, P. and Visser, M. (2004). A study of consumer behavior using laboratory data. *Experimental Economics*, 7(1):93–114.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Fisman, R., Jakiela, P., and Kariv, S. (2015a). How did distributional preferences change during the great recession? *Journal of Public Economics*, 128:84–95.
- Fisman, R., Jakiela, P., and Kariv, S. (2017). Distributional preferences and political behavior. *Journal of Public Economics*, 155:1–10.
- Fisman, R., Jakiela, P., Kariv, S., and Markovits, D. (2015b). The distributional preferences of an elite. *Science*, 349(6254):aab0096.
- Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5):1858–1876.

- Fligner, M. A. and Policello, G. E. (1981). Robust rank procedures for the behrens-fisher problem. Journal of the American Statistical Association, 76(373):162–168.
- Fong, C. M. and Luttmer, E. F. (2011). Do fairness and race matter in generosity? evidence from a nationally representative charity experiment. Journal of Public Economics, 95(5-6):372–394.
- Foucault, M. and Bono, P.-H. (2020). Enquête 2020 sur les maires de france. Technical report, CEVIPOF.
- Fréchette, G. R. (2016). Experimental economics across subject populations. The handbook of experimental economics, 2:435–480.
- Frigau, L., Medda, T., and Pelligra, V. (2019). From the field to the lab. an experiment on the representativeness of standard laboratory subjects. Journal of Behavioral and Experimental Economics, 78:160–169.
- Frijters, P., Kong, T. S., and Liu, E. M. (2015). Who is coming to the artefactual field experiment? participation bias among chinese rural migrants. Journal of Economic Behavior & Organization, 114:62–74.
- Gilovich, T., Griffin, D., and Kahneman, D. (2002). Heuristics and biases: The psychology of intuitive judgment. Cambridge university press.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. Journal of the Economic Science Association, 1(1):114–125.
- Gross, J. and Kaiser, D. (1996). Two simple algorithms for generating a subset of data consistent with warp and other binary relations. Journal of Business & Economic Statistics, 14(2):251–255.
- Gulzar, S. (2021). Who enters politics and why? Annual Review of Political Science, 24:253–275.
- Gustafsson, A. (2019). Busy doing nothing: why politicians implement inefficient policies. Constitutional Political Economy, 30(3):282–299.
- Hafner-Burton, E. M., Hughes, D. A., and Victor, D. G. (2013). The cognitive revolution and the political psychology of elite decision making. Perspectives on Politics, pages 368–386.

- Halevy, Y., Persitz, D., and Zrill, L. (2018). Parametric recoverability of preferences. Journal of Political Economy, 126(4):1558–1593.
- Harstad, R. M. and Selten, R. (2016). Diminished-dimensional political economy. European Economic Review, 83:213–219.
- Heufer, J. and Hjertstrand, P. (2015). Consistent subsets: Computationally feasible methods to compute the houtman–maks-index. Economics Letters, 128:87–89.
- Heufer, J. and Hjertstrand, P. (2019). Homothetic efficiency: Theory and applications. Journal of Business & Economic Statistics, 37(2):235–247.
- Hjort, J., Moreira, D., Rao, G., and Santini, J. F. (2021). How research affects policy: Experimental evidence from 2,150 brazilian municipalities. American Economic Review, 111(5):1442–80.
- Holtz-Eakin, D. and Rosen, H. S. (1989). The ‘rationality’ of municipal capital spending: Evidence from new jersey. Regional Science and Urban Economics, 19(3):517–536.
- Houtman, M. and Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. Kwantitatieve methoden, 19(1):89–104.
- Jacobs, A. M. (2011). Governing for the long term: Democracy and the politics of investment. Cambridge University Press.
- Jacquemet, N., James, A., Luchini, S., and Shogren, J. F. (2017). Referenda under oath. Environmental and Resource Economics, 67(3):479–504.
- Jacquemet, N., Joule, R.-V., Luchini, S., and Shogren, J. F. (2013). Preference elicitation under oath. Journal of Environmental Economics and Management, 65(1):110–132.
- Jacquemet, N., Luchini, S., Malezieux, A., and Shogren, J. F. (2020). Who’ll stop lying under oath? empirical evidence from tax evasion games. European Economic Review, 124:103369.
- Jennings, C. (2011). The good, the bad and the populist: A model of political agency with emotional voters. European Journal of Political Economy, 27(4):611–624.

- Jones, B. D. (2001). Politics and the architecture of choice: Bounded rationality and governance. University of Chicago Press.
- Kerschbamer, R. and Müller, D. (2020). Social preferences and political attitudes: An online experiment on a large heterogeneous sample. Journal of Public Economics, 182:104076.
- Kliemt, H. (2005). Public choice and political philosophy: Reflections on the works of gordon spinoza and david immanuel buchanan. Public Choice, 125(1):203–213.
- Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. Journal of Econometric Methods, 1(1):23–41.
- Laffont, J.-J. (2000). Incentives and political economy. Oxford University Press.
- Le Maux, B. (2009). How do policy-makers actually solve problems? evidence from the french local public sector. Economics & Politics, 21(2):201–231.
- Lee, N. (2021). Do policy makers listen to experts? evidence from a national survey of local and state policy makers. American Political Science Review, pages 1–12.
- LeVeck, B. L., Hughes, D. A., Fowler, J. H., Hafner-Burton, E., and Victor, D. G. (2014). The role of self-interest in elite bargaining. Proceedings of the National Academy of Sciences, 111(52):18536–18541.
- Linde, J. and Vis, B. (2017). Do politicians take risks like the rest of us? an experimental test of prospect theory under mps. Political Psychology, 38(1):101–117.
- Mansbridge, J. (2003). Rethinking representation. American Political Science Review, pages 515–528.
- Martinez-Bravo, M. (2014). The role of local officials in new democracies: Evidence from indonesia. American Economic Review, 104(4):1244–87.
- Martinez-Bravo, M. (2017). The local political economy effects of school construction in indonesia. American Economic Journal: Applied Economics, 9(2):256–89.
- Martinez-Bravo, M., Mukherjee, P., and Stegmann, A. (2017). The non-democratic roots of elite capture: Evidence from soeharto mayors in indonesia. Econometrica, 85(6):1991–2010.

- Mattei, A. (2000). Full-scale real tests of consumer behavior using experimental data. Journal of Economic Behavior & Organization, 43(4):487–497.
- Mattozzi, A. and Merlo, A. (2008). Political careers or career politicians? Journal of Public Economics, 92(3-4):597–608.
- Messner, M. and Polborn, M. K. (2004). Paying politicians. Journal of Public Economics, 88(12):2423–2445.
- Mueller, D. C. (2003). Public choice III. Cambridge University Press.
- Mullainathan, S., Schwartzstein, J., and Congdon, W. J. (2012). A reduced-form approach to behavioral public finance. Annal Review of Economics, 4(1):511–540.
- Müller, D. (2019). The anatomy of distributional preferences with group identity. Journal of Economic Behavior & Organization, 166:785–807.
- Muller, D. and Page, L. (2016). Born leaders: political selection and the relative age effect in the us congress. Journal of the Royal Statistical Society. Series A (Statistics in Society), pages 809–829.
- Murphy, J. H. and Banerjee, S. (2015). A caveat for the application of the critical cost efficiency index in induced budget experiments. Experimental Economics, 18(3):356–365.
- Nielsen, P. A. and Moynihan, D. P. (2017). How do politicians attribute bureaucratic responsibility for performance? negativity bias and interest group advocacy. Journal of Public Administration Research and Theory, 27(2):269–283.
- Osborne, M. J. and Slivinski, A. (1996). A model of political competition with citizen-candidates. The Quarterly Journal of Economics, 111(1):65–96.
- Palfrey, T. R. (2016). 6. experiments in political economy. In The Handbook of Experimental Economics, Volume 2, pages 347–434. Princeton University Press.
- Pedersen, R. T. and Dahlgaard, J. O. (2021). Political candidacy and sibling sex composition: Your sister will not make you run for office. Political Behavior, pages 1–19.
- Rachlinski, J. J. and Farina, C. R. (2001). Cognitive psychology and optimal government design. Cornell Law Review, 87:549.

- Salvino, R., Tasto, M. T., and Turnbull, G. K. (2012). A direct test of direct democracy: New england town meetings. Applied Economics, 44(18):2393–2402.
- Schnellenbach, J. and Schubert, C. (2015). Behavioral political economy: A survey. European Journal of Political Economy, 40:395–417.
- Schönhage, N. L. and Geys, B. (2021). Partisan bias in politicians’ perception of scandals. Party Politics, page 1354068821998024.
- Schreiber, D. (2017). Neuropolitics: Twenty years later. Politics and the Life Sciences, 36(2):114–131.
- Selten, R. (1991). Properties of a measure of predictive success. Mathematical Social Sciences, 21(2):153–167.
- Sheffer, L. and Loewen, P. J. (2019). Accountability, framing effects, and risk-seeking by elected representatives: An experimental study with american local politicians. Political Research Quarterly, 72(1):49–62.
- Sheffer, L., Loewen, P. J., Soroka, S., Walgrave, S., and Sheaffer, T. (2018). Non-representative representatives: An experimental study of the decision making of elected politicians. The American Political Science Review, 112(2):302–321.
- Simon, H. A. (1995). Rationality in political behavior. Political Psychology, pages 45–61.
- Sippel, R. (1997). An experiment on the pure theory of consumer’s behaviour. The Economic Journal, 107(444):1431–1444.
- Slonim, R., Wang, C., Garbarino, E., and Merrett, D. (2013). Opting-in: Participation bias in economic experiments. Journal of Economic Behavior & Organization, 90:43–70.
- Smeulders, B., Spieksma, F. C., Cherchye, L., and De Rock, B. (2014). Goodness-of-fit measures for revealed preference tests: Complexity results and algorithms. ACM Transactions on Economics and Computation (TEAC), 2(1):1–16.
- Stolwijk, S. and Vis, B. (2020). Politicians, the representativeness heuristic and decision-making biases. Political Behavior, pages 1–22.

- Thiemann, P., Schulz, J., Sunde, U., and Thöni, C. (2022). Selection into experiments: New evidence on the role of preferences, cognition, and recruitment protocols. Journal of Behavioral and Experimental Economics, 98:101871.
- Thomas, M. D. (2019). Reapplying behavioral symmetry: public choice and choice architecture. Public Choice, 180(1):11–25.
- Turnbull, G. K. and Chang, C. (1998). The median voter according to garp. Southern Economic Journal, pages 1001–1010.
- Turnbull, G. K. and Geon, G. (2006). Local government internal structure, external constraints and the median voter. Public Choice, 129(3-4):487–506.
- van Bruggen, P. and Heufer, J. (2017). Afriat in the lab. Journal of Economic Theory, 169:546–550.
- Varian, H. R. (1982). The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society, pages 945–973.
- Varian, H. R. (1988). Revealed preference with a subset of goods. Journal of Economic Theory, 46(1):179–185.
- Varian, H. R. (1990). Goodness-of-fit in optimizing models. Journal of Econometrics, 46(1-2):125–140.
- Varian, H. R. (1993). Goodness-of-fit for revealed preference tests. University of michigan ann arbor, Department of Economics.
- Vis, B. (2019). Heuristics and political elites’ judgment and decision-making. Political Studies Review, 17(1):41–52.
- Vis, B. and Stolwijk, S. (2020). Conducting quantitative studies with the participation of political elites: best practices for designing the study and soliciting the participation of political elites. Quality & Quantity, pages 1–37.
- Viscusi, W. K. and Gayer, T. (2015). Behavioral public choice: The behavioral paradox of government policy. Harvard Journal of Law & Public Policy, 38:973.
- Walgrave, S., Sevenans, J., Van Camp, K., and Loewen, P. (2018). What draws politicians’ attention? an experimental study of issue framing and its effect on individual political elites. Political Behavior, 40(3):547–569.

CHAPTER 1

Wallerstein, M. (2004). Behavioral economics and political economy. Nordic Journal of Political Economy, 30:37–48.

Wilson, R. K. (2011). The contribution of behavioral economics to political science. Annual Review of Political Science, 14:201–223.

Appendix A

Are Politicians Efficient Decision-Makers? A Revealed Preference Approach

Contents

A.1	Complements on experimental procedures	80
A.2	Complementary revealed preference analysis	97
A.3	Complementary regression analyses	114
A.4	Complementary power analyses	121
	References	128

A.1 Complements on experimental procedures

In this section, we further detail the practical aspects of our experiment and provide more information about the conduct of the sessions.

A.1.1 Instructions (translation from french)

Please do not read the instructions before being told so.

Likewise, please do not touch the computer before reading the instructions.

General Instructions :

Welcome

Thank your very much for attending this session.

Your dispose of instruction sheets at your station. These instructions will be read aloud. Please read them at the same time.

You are about to participate in an experiment related to economic decision-making. You are thus going to make a series of decisions. The possible reward you will receive at the end of the experiment will depend upon these decisions. Your decisions will remain anonymous and will not be circulated.

We ask you not to interact with other participants over the course of the experiment. We also ask you to turn off you cell phone. At the end of the experiment, we will ask you to sign a written commitment not to discuss the content of the experiment with other people who may participate later to this experiment.

This experiment was designed to last for approximately 1h30. However, take some time to think for every decision.

The remainder of the instructions presents the experiment's conduct.

The experiment is divided into five parts.

You will make decisions during the first three parts. The last part consists in a small questionnaire.

A reminder of your advancement within the experiment will be visible in the top-left corner of your screen.

At the beginning of each part, *Au début de chaque partie*, the instructions and decision rules relative to this part will be exposed.

Then, you will make your decisions.

Finally, you will go through the next part.

And so forth until the end of the experiment.

If you have any questions, do not hesitate to notify us by raising your hand. We will come answer to you individually. However, we kindly ask you to wait for the end of the instructions before asking your questions.

At the end of the experiment, you will be given the opportunity to obtain a reward.

Your reward is composed of two elements:

- a fixed amount of 6 euros, received by everybody, to thank you for participating.
- a variable amount which depends upon the decisions made during the experiment: your decisions, but also possibly the decisions made by other participants from your session.

More specifically, every part of the experiment will entail a possible reward. The variable amount corresponds to the sum of the rewards from the various parts.

Each participant's reward will be computed at the end of the experiment by the computer and will be then communicated to you personally.

APPENDIX A

Only a third of the participants from your session will effectively receive a reward with a variable amount. These participants will be randomly selected by the computer at the end of the experiment. The other participants will only be given the fixed amount.

When making your decisions, the amounts of money will be stated in ECU (Experimental Currency Units).

The conversion rule is the following: $1 \text{ ECU} = 0.07 \text{ euros}$. Consequently, 14 ECU are worth 1 euro.

Before proceeding with the first part, the computer will randomly select one identifier for each participant. It is the key element in the anonymization process.

Instructions Part 1:

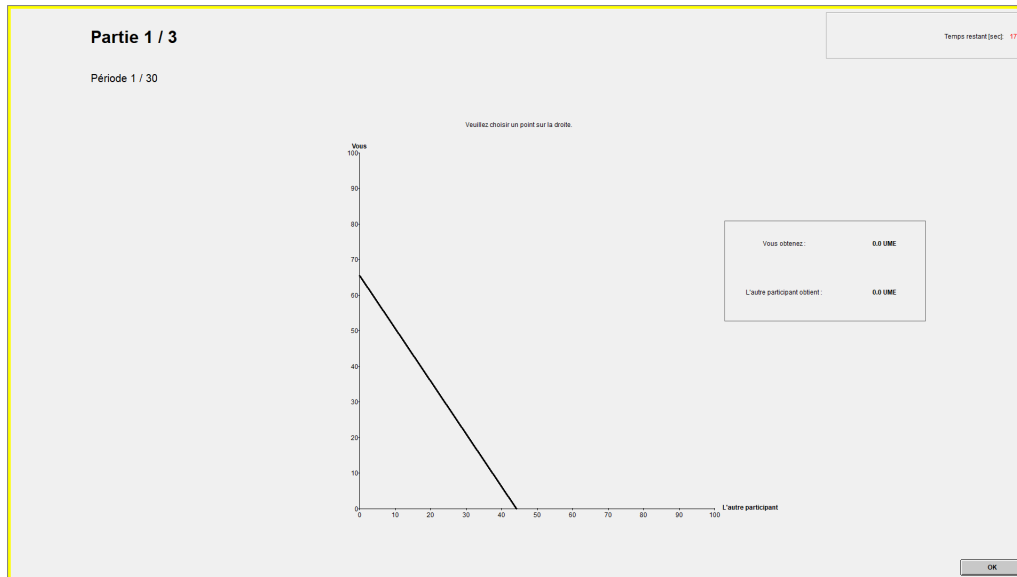
In the course of this part, you will make several decisions. From each decision will result a possible reward for both you and another participant in the room.

You will make 30 decisions during this part (hence 30 "periods").

For each period you will be matched with another participant in the room, randomly selected. Such participant changes across periods. You will never know the identity of any of the participants with whom you will be matched. Likewise, none of them will ever know your identity.

In every period, the screen will display a bold line in a mark. One example of screen is depicted in the figure below.

Every dot on the line represents an ECU allocation between yourself and the other participant. The amount that you will receive for every allocation is located on the y-axis (axis "You"). The amount received by the other participant for every allocation is located on the x-axis (axis "The other participant").



Your decision for this part consists in choosing a dot on the line by clicking on it.

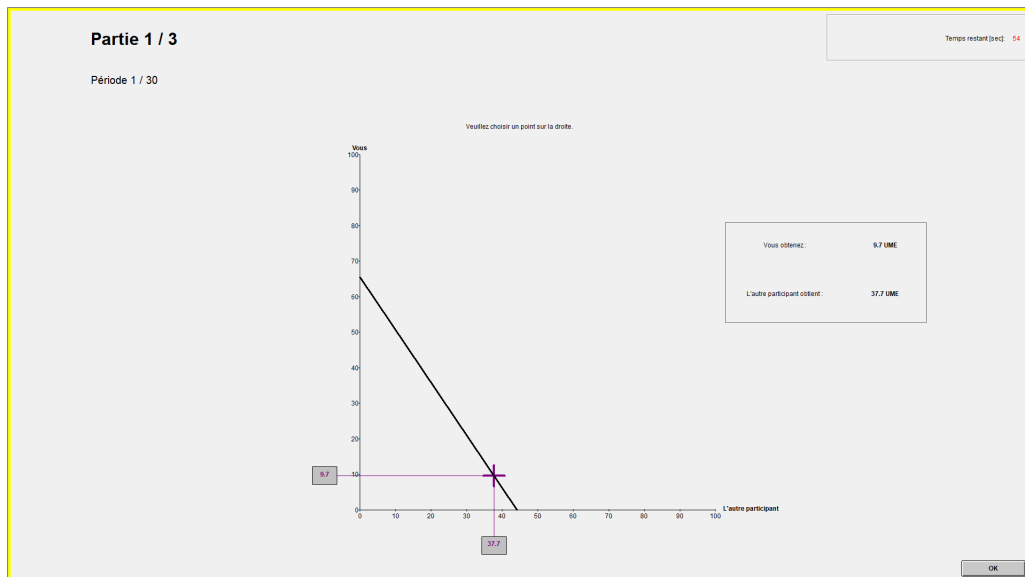
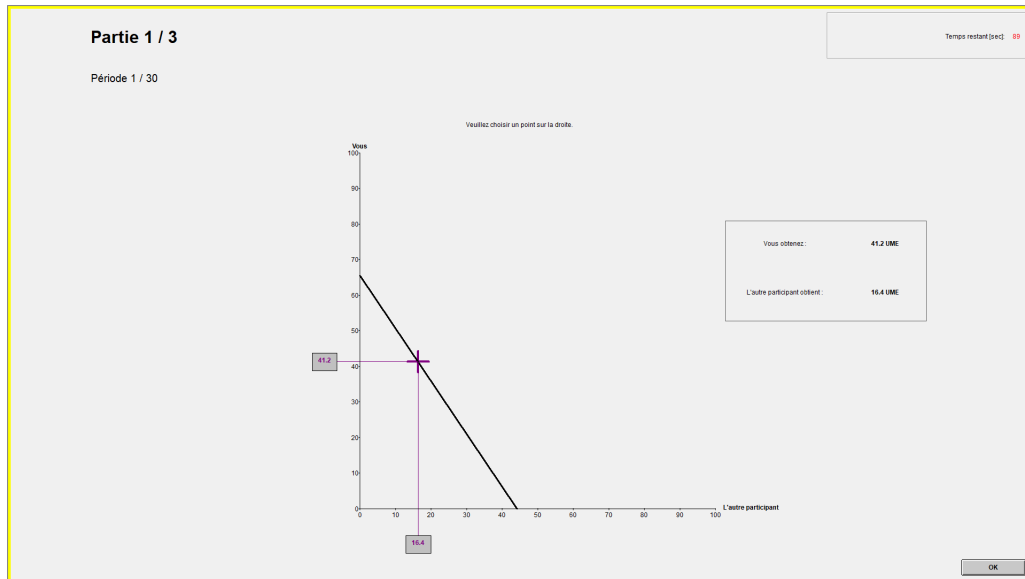
When you will click on a dot from the line, a purple cursor will appear and its position given each axis will be indicated in rectangles located under the x-axis and on the left of the y-axis. These rectangles thus also indicate how much ECU each participant will obtain if you choose this dot. A reminder of these information will be apparent in a box located on the right of the mark.

The two pictures in the next page introduce possible examples of dots on the line:

- in the first picture, you receive 41.2 ECU and the other participant receive 16.4 ECU.
- in the second picture, you receive 9.7 ECU and the other participant receive 37.7 ECU.

If you can elsewhere than on the line, the cursor will automatically be positionned on the line, on the nearest dot. Nevertheless, if you click elsewhere than in the zone defined by the boundaries of the mark, the cursos will be positionned at the intersection of the line and of one of the axes. For example, if you click on the left

APPENDIX A



of the axis "You" in one of the marks from the images on page 2, the cursors will be positionned on the dot for which you earn 65.1 ECU. Likewise, if you click below the axis "The other participant", the cursor will be positionned on the dot for which the other participant earns 43.8 ECU.

Once you have chosen a dot on the line, you will have to validate your choice by

clicking on "OK" on the bottom-right corner of the screen. You will then proceed with the next period.

In every period the line will change (as well as the participant with whom you will be matched). More precisely, the computer will randomly select one line among all possible lines that respect two requirements:

1. the line cannot cross one axis above 100 ECU.
2. the line cannot cross both axes below 50 ECU (but it is possible that it crosses one axis below 50 ECU, in which case it crosses the other axis above 50 ECU).

Remark: the draw is performed at the level of each participant, so that the line that is depicted on the screen is different across participants and across periods.

When computing your possible reward at the end of the experiment, the computer will proceed as follows. It will randomly select one period among the 30 periods. For this period, it will withhold what you chose for you and what the other participant with whom you were matched for this period chose for you. Your reward for this part corresponds to the sums of these two amounts.

Similarly, the possible reward for every participant with whom you were matched will depend on both what he chose for himself and what you chose for himself during the relevant period.

As a reminder, the conversion rule for this part is: 1 ECU = 0.07 euros.

The instructions for this part are now over. Please turn your attention to the screen.

A.1.2 Brief description of the subsequent experimental tasks

In addition to our main task exposed in sub-section 3.2, all subjects participated to several additional tasks whose content depends upon the sample under scrutiny.¹ Specifically, Students and Politicians from the Lab participated in additional distribution games, whereas Non-Students and Politicians from the Field participated in a modified version of the Trust Game. We will briefly summarize these additional tasks in this section.

Distribution games

The objective of this task was to drive the decision environment closer to the one that politicians deal with in their daily decisions. Specifically, the idea was to introduce two collective elements: that the decision is made for others and that it is reached through a collective process (a majority rule). In addition, the task allowed for different distributional motives as the main task, namely a difference between maximin and reducing inequalities, as well as a preference for efforts made by others.

Specifically, the task builds upon the allocation games proposed by Engelmann and Strobel (2004). Subjects had to choose between two options which allocated an ECU amount to a group of 3 subjects. The decision was repeated several times with changing the amounts in the two options so as to identify consistent patterns of choice. In particular, the objective was to disentangle a taste for efficiency, maximin or reducing inequalities.

Additionally, each member of the group was associated with a passive partner who did not participate in the decision. Passive partners had been exposed to an effort task prior to the experiment. In a given group, passive partners could then be ranked by their effort. Subjects were divided into two treatments: half of them received the information as to the ranking of passive partners, while such an information remained hidden to the other half.

We exposed all subjects to two different decision environments: under the veil and behind the veil of ignorance. In the first condition, subjects decided for another group and thus derived no personal gains from the decision. In the second condition,

¹As a reminder, all subjects faced our main task first and were not aware of the content of the other tasks when they were completing the first task. As a consequence, subjects' behavior in the main task cannot be affected by the other tasks.

subjects decided for their own group and thus derived a personal gain from the decision. Besides, in the second condition, subjects simultaneously decided as under a dictatorial system (each decision had one chance over three to be implemented) or under a majority system (the option with the highest number of votes within the group was implemented).

Trust Game

The objective of this task was to elicit Politicians' sensitivity to a norm of reciprocity. Indeed, politicians are regularly confronted to a situation where they have to decide upon reciprocating or not, starting with their relationship with the people that elect them.

In experimental economics, reciprocity is traditionally measured through the Trust Game. The game proceeds as follows: a subject called the "trustor" is given an amount of ECU to be divided with another subject, called the "trustee". Any ECU given to the trustee is multiplied by 3. Then, the trustee is asked to divide the tripled amount between himself and the trustor. The trustee's behavior (i.e., the amount given back to the trustor) is then considered as indicative of a willingness to reciprocate, that is to reward a positive behavior.

In order to isolate reciprocity from unconditional motives that may also explain why the trustee would send back some money (e.g., altruism or inequality-aversion), we use a dyadic design inspired by Cox (2004). The idea is to control for unconditional motives by comparing the standard Trust Game with a Dictator Game that shares the same structure. For instance, if in the Trust Game a trustee has to divide 180 ECU because the trustor sent him 60 ECU, then the Dictator Game would endow a dictator with 180 ECU to be divided with a passive recipient.

Besides, in order to control for an effect of initial inequalities induced by the amount sent by the trustor, we implement a strategy-method version of the game, as proposed by Di Bartolomeo and Papa (2016). Indeed, the inequalities are not the same when the trustor gives few ECU than when he gives everything, which may interact with the sensitivity to some motives. In other words, we ask the trustee about his behavior for all possible amounts that the trustor could give.

Additionally, our methodological contribution is to implement a within version

of the game where each subject is his own control. Specifically, all subjects play the Trust Game and the identically-structured Dictator Game, in random order.² Ultimately, we are capable of computing a reciprocity score for each subject and each initial situation as the difference between what he sent back as a trustee and what he gave as a dictator.

²For the Dictator Game, subjects are told that the computer will randomly select one endowment to be played with within a pre-determined set. The pre-determined set corresponds to all amounts that are available for a trustor to give in the Trust Game. Then, the strategy method is used to obtain the amount that subjects give for each possible endowment.

A.1.3 Questionnaire and definition of the socio-demographic variables

The end-of-experiment questionnaire included several socio-demographic questions from which we constructed our control variables. The wording of each question is presented in this section, along with the definition and type of the corresponding variable. We also specify when the question/variable was specific to some experimental groups or if it was altered across groups.

Gender You are:

- A woman
- A man

Variable: dummy.

Education (*italics identify categories not included in the questionnaire for Politicians from the Lab*) Your highest degree (or equivalent) is:

- General certificate secondary education
- A-level
- *First-year university degree*
- *Second-year university degree*
- Third-year university degree (academic license)
- *Master 1*
- Master 2 (master's degree)
- *Above Master*
- Other

Variable: dummy on whether the subject holds a higher education degree (\geq third-year university).

Age How old are you?

Variable: continuous.

Field of Studies (*italics identify categories included only in the questionnaires for Politicians from the Field and Non-Students*) What major did you study most during your education?

- Economics
- Economic and Social Administration (ESA)
- Law
- Management
- Political Science
- Literature/*Philosophy*
- Chemistry
- Physics
- Medicine
- *Sociology*
- *History/Geography*
- *Modern Language*
- *Mathematics*
- *Computer Science*
- *Engineering Science*
- *Biology/Geology*
- *Arts*
- Other

Variable: categorical with four categories:

1. Fundamental sciences: Chemistry, Medicine, Physics, *Mathematics, Computer Science, Engineering Science, Biology/Geology*
2. Economics: Economics, ESA, Management
3. Humanities: Law, Literature, Political Science, *Arts, Modern Language, Sociology, History/Geography*
4. Other

Occupation Do you have an occupation?

- Private-sector worker
- Public-sector worker
- Self-employed
- Retired
- Unemployed
- Inactive

Variable: categorical variable with the same categories, except for merging "Inactive" with "Unemployed" (hence 5 remaining categories).

Income In which monthly income group are you?

- < 1300€
- Between 1300 and 2000€
- Between 2000 and 3000€
- > 3000€
- Do not know
- Do not want to answer

Variable: categorical variable with the same categories, except for merging "< 1300€" with "Between 1300 and 2000€" and for merging "Do not know" with "Do not want to answer" (hence 4 remaining categories).

Current elected status at the local level (*Politicians only*) In your town for the current period (2014-2020), you are:

- Mayor
- Vice-Mayor
- Local councillor

Variable: categorical variable with the same categories.

Current elected status at other administrative layers (*Politicians only*) In addition, for the current period (2014-2020), are you:

- President of an intermunicipal structure
- Vice-President of an intermunicipal structure
- Councillor in an intermunicipal structure
- County councillor (Département)
- District councillor (Région)
- Member of Parliament
- Senator
- Other

Variable: dummy on whether the subject holds a term at a higher administrative layer in addition to his local term.³

Political experience (number of terms) (*Politicians only*) How many terms have you held in your town (current term included):

- 1
- 2
- 3
- 4 or more

Variable: categorical variable with the same categories.

³In our dataset, considering that only 6 subjects are county councillors and 2 subjects are district councillors, the majority of "higher-terms" consists of intermunicipal terms.

Pre-2014 elected status (*Politicians only*) Before 2014, have you been:

- Mayor
- Vice-Mayor
- Municipal councillor
- President of an intermunicipal structure
- Vice-President of an intermunicipal structure
- Councillor in an intermunicipal structure
- County councillor (Département)
- District councillor (Région)
- Member of Parliament
- Senator
- Other

Political experience (number of years) (*Politicians only*) Overall, how many years have you been elected?

Variable: continuous variable.

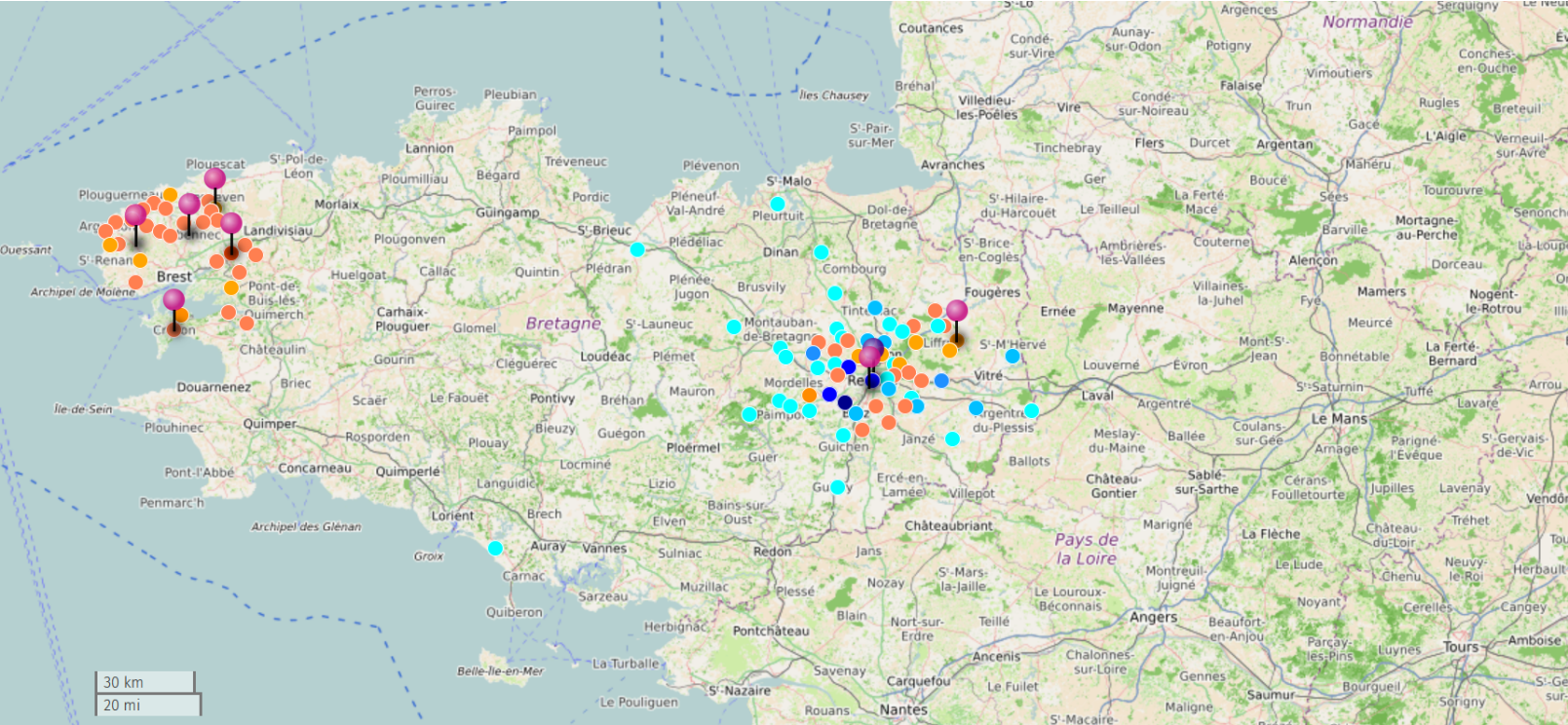
A.1.4 Sessions' information

Table A.1: Experimental sessions

Session	Day of week	Date	Hour	Group of subjects	Type of session	Number of subjects	Average payoff (€)	Maximum payoff (€)
1	Fri	10/26/2018	10:15-12:15	Politicians	Lab	6	17.21	39.89
3	Tue	10/30/2018	10:15-12:15	Politicians	Lab	6	15.96	35.90
5	Wed	10/31/2018	10:15-12:15	Politicians	Lab	6	16.47	27.82
7	Thu	11/08/2018	10:15-12:15	Politicians	Lab	9	15.56	35.43
9	Wed	11/14/2018	12:00-14:00	Politicians	Lab	6	15.96	36.69
10	Wed	11/14/2018	18:00-20:00	Politicians	Lab	6	17.46	40.85
11	Thu	11/15/2018	10:15-12:15	Politicians	Lab	9	15.30	32.44
13	Thu	11/22/2018	18:00-20:00	Politicians	Lab	3	17.25	37.75
14	Fri	11/23/2018	14:00-16:00	Politicians	Lab	9	13.57	28.34
15	Tue	11/27/2018	13:00-15:00	Politicians	Lab	6	17.16	39.21
16	Wed	11/28/2018	10:15-12:15	Politicians	Lab	3	14.17	28.52
2	Fri	10/26/2018	14:00-16:00	Student Control	Lab	21	16.02	42.33
4	Tue	10/30/2018	14:00-16:00	Student Control	Lab	21	16.16	37.91
6	Wed	10/31/2018	14:00-16:00	Student Control	Lab	24	16.03	38.32
8	Thu	11/08/2018	14:00-16:00	Student Control	Lab	24	15.04	38.41
12	Thu	11/15/2018	14:00-16:00	Student Control	Lab	21	16.17	35.48
17	Thu	01/17/2019	14:00-16:00	Student Control	Lab	12	17.18	43.21
18	Tue	09/10/2019	18:00-19:30	Politicians	Field	10	13.51	36.29
19	Wed	09/11/2019	17:00-18:30	Politicians	Field	9	11.75	26.23
20	Wed	09/11/2019	18:30-20:00	Politicians	Field	11	13.29	47.79
21	Wed	09/25/2019	17:00-18:30	Politicians	Field	10	14.12	38.10
22	Mon	10/14/2019	18:30-20:00	Politicians	Field	10	14.93	46.05
23	Thu	11/28/2019	18:00-19:30	Politicians	Field	10	12.00	29.72
24	Tue	12/03/2019	18:00-19:30	Politicians	Field	6	14.90	32.93
25	Wed	01/22/2020	18:00-19:30	Politicians	Field	6	15.70	43.54
26	Sat	01/25/2020	10:00-11:30	Non-student Control	Lab	13	12.35	37.03
27	Sat	01/25/2020	14:00-15:30	Non-student Control	Lab	12	12.59	34.69
28	Sat	01/25/2020	16:30-18:00	Non-student Control	Lab	7	14.59	38.70
29	Wed	01/29/2020	18:00-19:30	Non-student Control	Lab	22	13.65	42.66
30	Sat	02/01/2020	10:00-11:30	Non-student Control	Lab	16	14.33	42.30
31	Sat	02/01/2020	14:00-15:30	Non-student Control	Lab	14	13.83	47.95
32	Sat	02/01/2020	16:30-18:00	Non-student Control	Lab	5	10.02	26.11

Note: Average/maximum payoff is computed over the entire experiment, not only on the game presented here. Indeed the experiment included other games whose nature varied across the two experimental waves. Payoffs thus vary across the four groups of subjects since the last two experienced shorter sessions (by 30 minutes) and thus have lower average payoffs despite the fact that they earn approximately the same in the game under scrutiny. The minimum payoff is always 6 euros, for the two thirds subjects who did not get selected for payment.

Figure A.1: Map of Politicians' towns



Note: Each dot represents a town. The darker the dot's color, the bigger the number of subjects coming from the corresponding town (maximum is 6). Lab sessions are painted in blue whereas field sessions are painted in orange. The pins identify the towns where the experiment took place. The violet one corresponds to the University of Rennes 1. The seven others correspond to the towns that hosted the field sessions.

Source: This map was produced with uMap (<https://umap.openstreetmap.fr>), which freely allows to personalize maps from the open-source collaborative project OpenStreetMap.

Figure A.2: Pictures of some experimental rooms



(a) Lab
(LABEX-EM, University of Rennes 1)



(b) Field
(Rennes Métropole)



(c) Field
(Liffré-Cormier Communauté)



(d) Field
(CC du Pays des Abers)



(e) Field
(Communauté Lesneven Côte des Légendes)



(f) Field
(CC Presqu'île de Crozon-Aulne maritime)

A.2 Complementary revealed preference analysis

This section exposes how we computed each of our three rationality indices: the Afriat's Efficiency Index (AEI), the Houtman and Maks's Index (HMI) and the Varian's Efficiency Vector (VEV). We also provide data description for the last two indices as a complement to the analysis presented in section 3. Then, we expose the evolution of the inconsistencies over the course of the experiment. Finally, we evaluate the importance of a possible limitation with the use of the AEI (and by extension the VEV) to compare the rationality across groups in our dataset: its inability to rule-out all types of violations.

A.2.1 Computation of the indices

A.2.1.1 The Afriat (1972, 1973)'s Efficiency Index (AEI)

The AEI corresponding to subject i 's dataset can be constructed following Afriat (1973, 1987) (see also Murphy and Banerjee, 2015):

First, for any couple of observations π_s^i and π_s^j , let:

$$\begin{aligned} D_i^{j,k} &= p_i^j \pi_i^k - 1 \\ d_i^{j,k} &= \max(D_i^{j,k}, D_i^{k,j}) \end{aligned}$$

Second, define the cross cost efficiency index between the two observations as:

$$e_i^{j,k} = 1 - \max(0, -d_i^{j,k})$$

Finally, the AEI over the entire dataset is given by:

$$e_s^* = \min_{\substack{\forall j,k \\ j \neq k}} (e_i^{j,k})$$

A.2.1.2 The Houtman and Maks (1985)'s Index (HMI)

In order to compute the HMI, we resort to the second algorithm proposed by Gross and Kaiser (1996). Such algorithm is all the more efficient as it does not make use of any optimization routine. However, it is limited to two-dimensional commodity spaces. For a more general algorithm which supports commodity spaces of higher dimensions, see the second algorithm in Heufer and Hjertstrand (2015). Although the algorithm was originally designed to test for WARP violations, it naturally extends to GARP (see Heufer and Hjertstrand, 2015). The overall principle is to start from the subject i 's dataset and remove observations until the remaining observations satisfy GARP. Any removed observation will then be attributed a 0 at its corresponding coordinate in the vector \tilde{e}_i , while any remaining observation will be attributed a 1. The HMI is finally computed as:

$$\text{HMI} = \frac{1}{30} \sum_{j=1}^{30} e_i^j$$

More specifically, Gross and Kaiser (1996) took a graph-theoretic approach wherein every observation is interpreted as a node of a graph. When two observations constitute a violation of GARP, they are said to be "adjacent", which means that there exists an edge connecting them. As a consequence observations that do not take part in a GARP-violation are simply nodes with empty space surrounding them. The number of nodes to which a given node is adjacent is called the "degree" of such a node. Let A_i^j the set of nodes adjacent to allocation π_i^j and $1A_i^j$ the set of nodes adjacent to allocation π_i^j and with degree 1.

Based on that setup, the algorithm runs through two-stages, until no allocation is removed anymore:

1. whenever $\text{degr}(\pi_i^j) = \max(\text{degr}(\pi_i^1), \dots, \text{degr}(\pi_i^{30}))$ and $\text{degr}(\pi_i^j) > \text{degr}(\pi_i^k)$ $\forall \pi_i^k \in A_i^j$, remove π_i^j .
2. whenever $\text{degr}(\pi_i^j) = \max(\text{degr}(\pi_i^1), \dots, \text{degr}(\pi_i^{30})) = \text{degr}(\pi_i^k)$ for some $\pi_i^k \in A_i^j$:
 - (a) if $1A_i^j \neq \emptyset$, remove π_i^j .
 - (b) if $1A_i^k \neq \emptyset$, remove π_i^k .
 - (c) if $1A_i^j = 1A_i^k = \emptyset$, randomly remove either π_i^j or π_i^k .

A.2.1.3 The Varian (1990, 1993)'s Efficiency Vector (VEV)

As a reminder, Afriat (1967)'s theorem shows an equivalence between the following conditions for a given finite set of j observations (Varian, 1982):⁴

- (i) there exists a well-behaved utility function $u(\cdot)$ that rationalizes the set of observations: we have $u(\pi_i^j) \geq u(\pi_i^k)$ whenever $p_i^j \pi_i^j \geq p_i^j \pi_i^k$.
- (ii) the set of observations satisfies GARP: we cannot simultaneously have $p_i^j \pi_i^j \geq p_i^j \pi_i^k$ and $p_i^k \pi_i^k > p_i^k \pi_i^j$ for all observations j and k .
- (iii) there exists j numbers U_i^j complying with the following inequalities:

$$U_i^j \geq U_i^k \quad \text{whenever} \quad p_i^j \pi_i^j \geq p_i^j \pi_i^k$$

$$U_i^j > U_i^k \quad \text{whenever} \quad p_i^j \pi_i^j > p_i^j \pi_i^k$$

Bulding upon the notion of e -rationalization introduced by Halevy et al. (2018) (also see Heufer and Hjertstrand, 2019), such a theorem incorporates the "relaxed" version of GARP as:

- (i) there exists a well-behaved utility function $u(\cdot)$ that e -rationalizes the set of observations: we have $u(\pi_i^j) \geq u(\pi_i^k)$ whenever $e_i^j p_i^j \pi_i^j \geq p_i^j \pi_i^k$.
- (ii) the set of observations satisfies GARP(\tilde{e}_i): we cannot simultaneously have $e_i^j p_i^j \pi_i^j \geq p_i^j \pi_i^k$ and $e_i^k p_i^k \pi_i^k > p_i^k \pi_i^j$ for all observations j and k .
- (iii) there exists j numbers U_i^j complying with the following inequalities:

$$U_i^j \geq U_i^k \quad \text{whenever} \quad e_i^j p_i^j \pi_i^j \geq p_i^j \pi_i^k \quad (\text{A.2a})$$

$$U_i^j > U_i^k \quad \text{whenever} \quad e_i^j p_i^j \pi_i^j > p_i^j \pi_i^k \quad (\text{A.2b})$$

⁴Such a formulation of the theorem omits the Afriat's inequalities condition. Furthermore, the condition (iii) is a standard reformulation of GARP that is common in producer analysis (Varian, 1984). Its interpretation is simply that if subject i places a higher value on allocation π_i^j than on allocation π_i^k , then for subject i to be a utility-maximizer it should not be the case that π_i^j was strictly affordable when π_i^k was chosen (i.e., it should cost at least as much, or more, depending on the stringency of the inequality). Also see Cherchye et al. (2015).

Solving the problem defined by (A.2a) and (A.2b) provides a way of computing the Varian's Efficiency Vector (VEV). We follow Heufer and Hjertstrand (2019) and define the problem in the form of a Mixed-Integer Linear Programming (MILP) model.⁵ Basically, it amounts to transforming the problem into a system of linear expressions and to solve it using optimization routines.

Specifically, the model is formulated as a system with four types of constraints. The inequalities (A.2a) and (A.2b) are equivalent to the existence of j numbers $U_i^j \in [0, 1]$ ⁶, j numbers $e_i^j \in (0, 1]$, and $j(j - 1)$ integer numbers $X_s^{j,k} \in \{0, 1\}$ such that $\forall j, k$:

$$U_i^j - U_i^k < X_s^{j,k} \quad (c.1)$$

$$U_i^k - U_i^j \leq 1 - X_i^{j,k} \quad (c.2)$$

$$e_i^j p_i^j \pi_i^j - p_i^j \pi_i^k < A_i^j X_i^{j,k} \quad (c.3)$$

$$e_i^k p_i^k \pi_i^k - p_i^k \pi_i^j \leq A_i^j (1 - X_i^{j,k}) \quad (c.4)$$

with $A_i^j > p_i^j \pi_i^j$ a fixed number (set to 100 in our procedures).

The interpretation behind the integer numbers $X_s^{j,k}$ is that each of them should be equal to one (resp. 0) only when $U_i^j \geq U_i^k$ (resp. $U_i^j < U_i^k$). This requirement is ensured by the constraints (c.1) and (c.2). The integer numbers are then used to set up the constraints (c.3) and (c.4) on numbers e_i^j to meet with GARP(\tilde{e}_i) requirement that $\pi_i^j R_D(\tilde{e}_i) \pi_i^k \Rightarrow \neg \pi_i^k P_D(\tilde{e}_i) \pi_i^j$. Indeed, if we have $U_i^j \geq U_i^k$ (that is $X_i^{j,k} = 1$), then we must also have $e_i^k p_i^k \pi_i^k \leq p_i^k \pi_i^j$. That is because by (A.2a) we must have $\pi_i^j R_D(\tilde{e}_i) \pi_i^k$ ($e_i^j p_i^j \pi_i^j \geq p_i^j \pi_i^k$), which is not compatible with GARP(\tilde{e}_i) if we also have $\pi_i^k P_D(\tilde{e}_i) \pi_i^j$ ($e_i^k p_i^k \pi_i^k > p_i^k \pi_i^j$). Conversely, if we have $U_i^j < U_i^k$ (that is $X_s^{j,k} = 0$), then we must also have $e_i^j p_i^j \pi_i^j < p_i^j \pi_i^k$. That is because by (A.2b) we must have $\pi_i^k P_D(\tilde{e}_i) \pi_i^j$ ($e_i^k p_i^k \pi_i^k > p_i^k \pi_i^j$), which is not compatible with GARP(\tilde{e}_i) if we also have $\pi_i^j R_D(\tilde{e}_i) \pi_i^k$ ($e_i^j p_i^j \pi_i^j \geq p_i^j \pi_i^k$).

⁵MILP models are simply special cases of standard Linear Programming models where some variables are restricted to be integers. Such class of models has been introduced to the revealed preference literature by Cherchye et al. (2008).

⁶The restriction on numbers U_i^j to the unit interval is harmless as rescaling will not change the underlying revealed preference conditions.

Finally, Heufer and Hjertstrand (2019) propose to calculate the vector $\tilde{\mathbf{e}}_i$ by solving the following minimization problem subject to the above constraints:⁷

$$\min \left\{ \sum_{j=1}^n (1 - e_i^j) \right\}$$

The MILP formulation is convenient for several reasons. First, such a problem always have an exact solution and any local solution is also a global solution. We are thus guaranteed to find a global optimum for the VEV. Second, the solution to our MILP problem will satisfy both necessary and sufficient conditions: the vector we will find will be a VEV and there will exist no "better-fit" vector (i.e., an identical vector except for strict improvements in the form of lower required adjustments in some coordinates). Such a condition is generally not met with existing iterative algorithms (as in the original proposal by Varian, 1990, 1993), which would identify budget adjustments that are not necessarily minimal due to the computational complexity of the problem (NP-hard). Third, MILP models are commonly accepted as a way to deal with NP-hard problems and there now exists extremely powerful and well-documented solvers able to handle sizeable problems (in our case the model for each subject is composed of 960 variables and 5400 linear inequality constraints).⁸

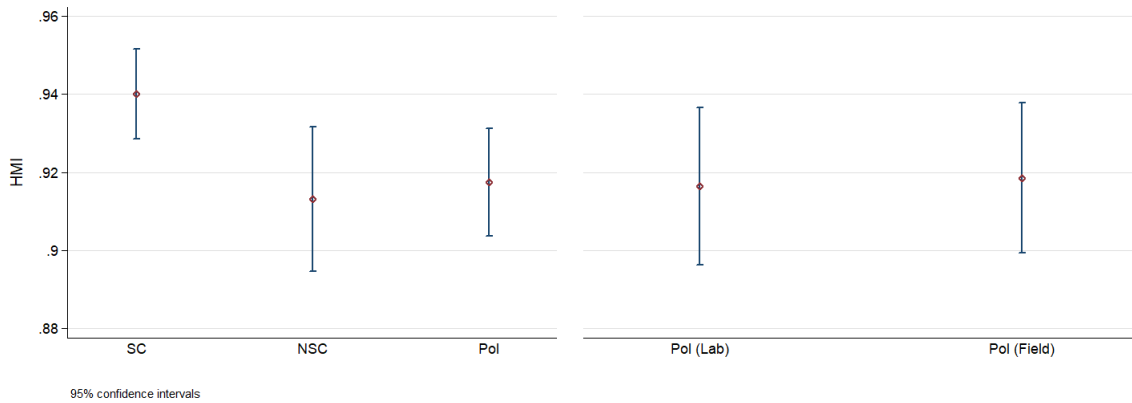
⁷Such an objective function implies a specific "norm of closeness" between GARP violations and strict rationality (that is a vector filled with 1), namely the taxicab norm (also known as "L1-norm"). Other candidates for such a norm have been proposed in the literature: the Euclidian norm or "L2-norm" (Varian (1990) and Alcantud et al. (2010)), the maximum norm (Varian (1993) and Cox (1997)), or the p-norm which generalizes all the previously-mentioned norms (Smeulders et al. (2014), also mentioned in Heufer and Hjertstrand (2019)). Comparing the results across all these norms is beyond the scope of the present work and we simply decided to stick to Heufer and Hjertstrand (2019)'s methodology.

⁸Our results are obtained using version 12.10.0 of the IBM ILOG CPLEX Optimizer that can be freely downloaded thanks to IBM's Academic Initiative. We use the Open Programming Language (OPL) which is directly implementable within the software's build-in interface.

A.2.2 Additional results on between-groups comparisons

In this section, we report results on between-groups comparisons based on the HMI and on the VEV. The figure A.3 plots the mean HMI along with the 95% confidence interval for all the experimental groups. The table A.2 summarizes the information regarding the VEV.

Figure A.3: Houtman and Maks’s Index by groups



(a) Main groups

(b) Politicians sub-groups

One issue with the results obtained from solving the MILP is to find a way to efficiently summarize the information contained within each individual vector in a meaningful way (a problem that does not arise with single numbers like the AEI or the HMI). The table A.2 presents several descriptive statistics concerning the VEV akin to Heufer and Hjertstrand (2019)’s presentation⁹. In the table A.2a, the values for each group correspond to average values of each considered statistic computed at the individual level. In the table A.2b, the values for each group correspond to median values of each considered statistic computed at the individual level. For instance, the minimum values rely on computing $\min\{e_i^1, \dots, e_i^{30}\}$ for each subject i and averaging these values over all subjects in the corresponding group (table A.2a)

⁹Due to the randomization of budget line that sometimes allow for flat lines along an axis, one subject picked the origin of the mark as an allocation. This leads to a AEI of 0 and also causes convergence issues for the computation of the VEV because the e_i^j are not allowed to be null (see the definition of the MILP problem). This subject is then excluded from the analysis.

or computing the median of these values over all subjects in the corresponding group table A.2b).

Table A.2: Summary of Varian's Efficiency Vector across groups

(a) Mean summary						
	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
Mean	0.987	0.992	0.990	0.988	0.992	0.990
Minimum	0.849	0.898	0.874	0.823	0.927	0.880
First Quartile	0.991	0.995	0.993	0.995	0.992	0.993
Median	0.998	1.000	0.999	1.000	0.998	0.999
Third Quartile	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
(b) Median summary						
	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
Mean	0.998	0.998	0.998	0.997	1.000	0.999
Minimum	0.968	0.961	0.963	0.945	0.990	0.974
First Quartile	1.000	1.000	1.000	1.000	1.000	1.000
Median	1.000	1.000	1.000	1.000	1.000	1.000
Third Quartile	1.000	1.000	1.000	1.000	1.000	1.000
Maximum	1.000	1.000	1.000	1.000	1.000	1.000

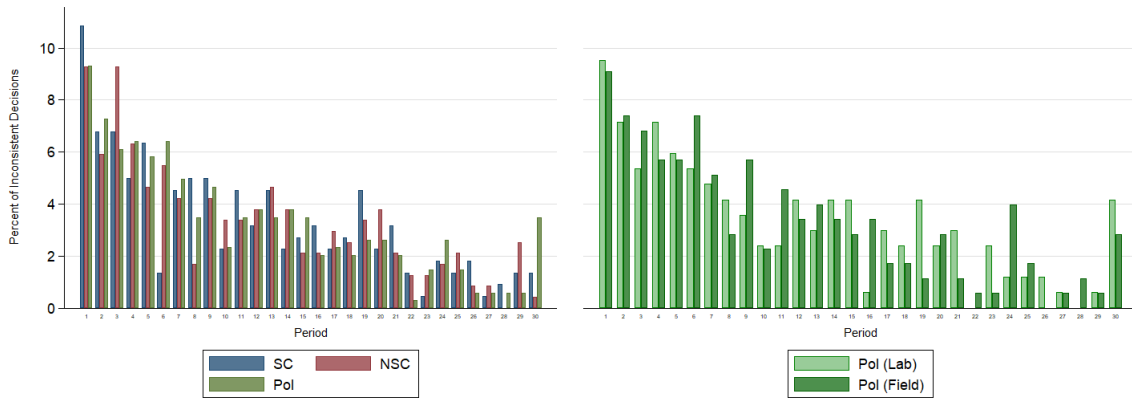
Note: The values for each group correspond to the average or median values of each considered statistic computed at the individual level. For instance, in table A.2a the minimum values rely on computing $\min\{e_i^1, \dots, e_i^{30}\}$ for each subject i and averaging these values over all subjects in the corresponding group.

A.2.3 Inconsistencies over time

In this section, we briefly describe the evolution of GARP violations over time for each experimental group. Indeed, both the HMI and the VEV provide information as to when our subjects deviate from GARP.

The algorithm we use to compute the HMI identifies the maximal set of observations that would pass the GARP test for a given subject. Consequently, we are able to identify the periods during which each subject violate GARP, though we are not capable of stating the extent of each violation. The figure A.4 displays the proportion of inconsistent decisions happening in each of the 30 periods in our experiment.

Figure A.4: Distribution of GARP-inconsistencies over time by groups



(a) Main groups

(b) Politicians sub-groups

Note: Each bar indicates the proportion of GARP-violations that falls into the corresponding period for the group under scrutiny.

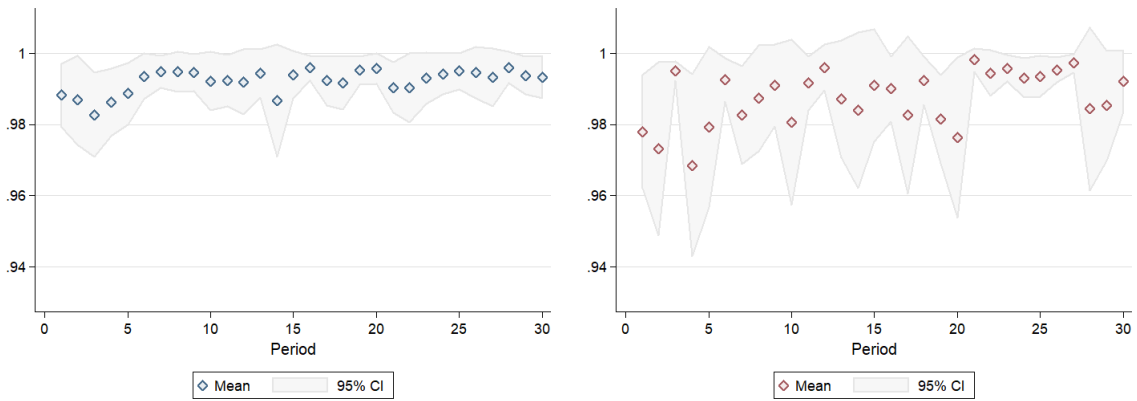
Overall, we find that choice inconsistencies occur at the beginning of the experiment: out of the 802 inconsistencies that we observe, 443 (55.24%) occur before the 10th period and 283 (35.29%) occur before the 5th period, against 132 (16.46%) after the 20th period and 57 (7.11%) occur after the 25th period. Although there

exists some visual differences across our groups¹⁰, this general pattern prevails in all groups. Such a finding is consistent with the idea that subjects may use the first rounds to get acquainted with the decision environment or may require few periods to realize exactly what they want to do.

In addition, the figures A.5 and A.6 plot the average required budget adjustment as well as its 95% confidence interval associated with each of the 30 budget sets for all our experimental groups. Such figures allow to further investigate whether the biggest violations also happen at the beginning of the experiment. Overall, we observe a slight decrease of budget waste over time, but the correlation is very close to zero (about 0.04). Subjects thus seem to deviate more often from GARP at the beginning of the experiment, but the extent of the violations is relatively constant over time on average.

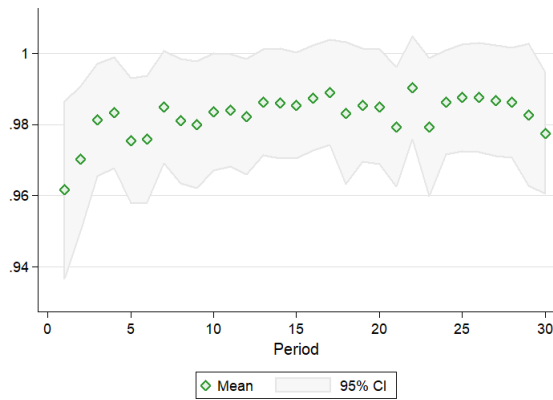
¹⁰All tests (t-test, FP test and ES test) fail to reject the null hypothesis of no difference between Politicians and any of the control groups or between Politicians from the Lab and Politicians from the Field (results unreported here).

Figure A.5: Distribution of budget adjustments (VEV) over time by groups



(a) Student Control

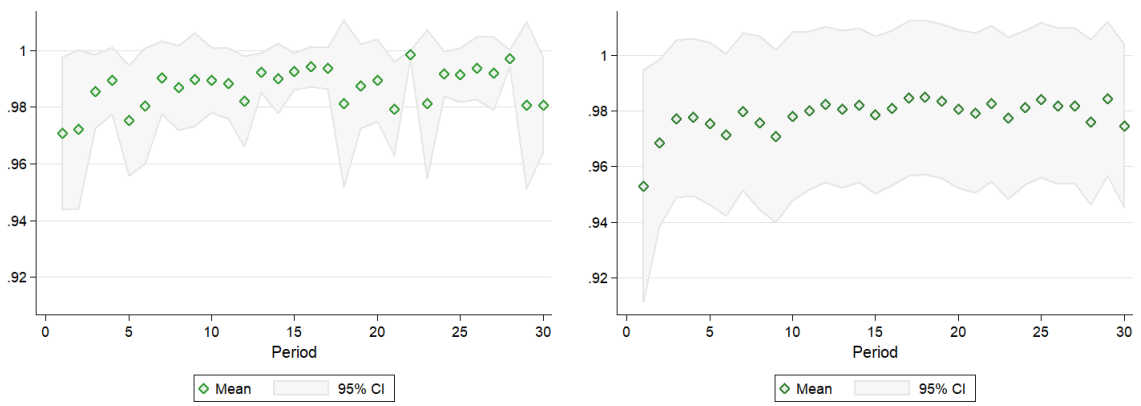
(b) Non-student Control



(c) Politicians

Note: Each dot is the average of the j -th element in the Varian's Efficiency Vector (VEV), with $j \in \llbracket 1, 30 \rrbracket$.
The shaded area correspond to the 95% interval around this average.

Figure A.6: Distribution of budget adjustments (VEV) over time by Politicians sub-groups



(a) Politicians (Lab)

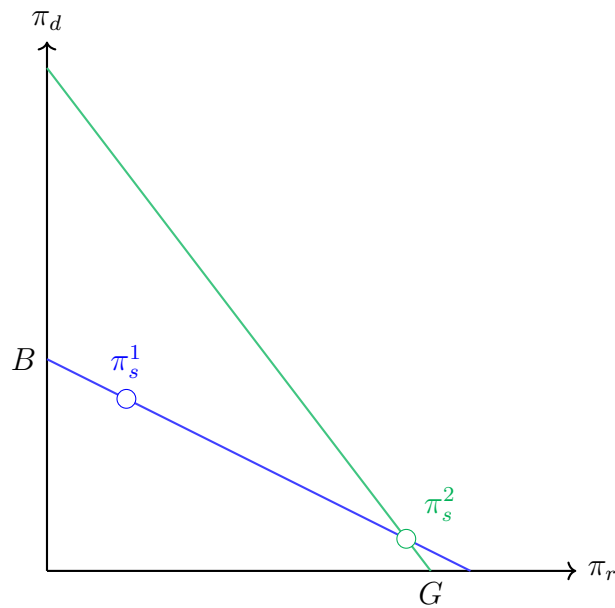
(b) Politicians (Field)

Note: Each dot is the average of the j -th element in the Variance Efficiency Vector (VEV), with $j \in \llbracket 1, 30 \rrbracket$.
The shaded area correspond to the 95% interval around this average.

A.2.4 Caveat with the use of AEI/VEV

A subtle point must be made concerning the case where $AEI = 1$, which corresponds to a situation where no perturbation of the budget constraints is needed to make the subject consistent with GARP. However, such a formulation is slightly misleading. Indeed it may happen that one GARP-inconsistent subject is attributed an AEI equal to 1. Such a case may happen when one allocation lies at the intersection of the two budget lines. An example of this situation is presented in the figure A.7: we have a GARP violation since $\pi_s^1 R_D \pi_s^3$ is contradicted by $\pi_s^3 P_D \pi_s^1$, but at the same time an infinitesimal shift in the budget constraint through π_s^3 would make π_s^3 unaffordable at prices p_s^1 and would remove the violation. As a consequence, we have $D_s^{1,3} = 0$ and $AEI = 1$. A similar argument can be made of subjects for which all elements in the VEV are equal to 1.

Figure A.7: CEV example



Such a violation of GARP that is undetected by the AEI has been termed a "cost-efficient violation" (CEV) by Murphy and Banerjee (2015) (though already pinpointed by Andreoni and Miller, 2002): it is efficient as it does not involve any

waste in expenditure, but it still violates GARP. It is opposed to the "cost-inefficient violation" (CIV) that is presented in most textbooks (see figure 1.2 in the chapter 1).¹¹

As emphasized by Murphy and Banerjee (2015), not accounting for the incapacity of the AEI to discriminate between truly GARP-consistent subjects and CEV-type violators might bias the results. Specifically, the authors highlight a tradeoff between increasing Bronars (1987)'s power (which implies reducing the angle between intersecting budget lines) and decreasing the likelihood of CEV (which implies raising the angle). Such tradeoff is all the more important that in continuous budget environments, the experimenter cannot preclude the possibility that some of subjects' choices lie at the intersection of budget lines. Such a preclusion would not even be desirable as exposing subjects to a very large variety of budget situations is a highly-valuable feature of the experiment. Furthermore, this tradeoff may depend upon the subject pool and consequently introduce some differences in the measurement of rationality across our experimental groups.

What is the importance of CEV in our data? Ultimately, only a minority of our subjects show CEV violations: overall CEV represent only 2.13% of violations and only 44 out of 353 subjects exhibit at least one CEV (hence 12.75%). Among these, only 9 shows no CIV (2.55% of total subjects) and are thus undetected as GARP violators with the AEI or the VEV. As a consequence, we deem the existence of CEV sufficiently small to have little impact on our results. We also observe that among the 9 subjects with only CEV, there are a majority of Politicians: 2 from the Lab and 5 from the Field, against 1 Student and 1 Non-Student. Such an observation is compatible with the fact that we observe very similar proportions of GARP-consistent subjects between Politicians and Non-Students, while Politicians (especially from the Field) appear slightly closer to rationality based on the AEI or the VEV than Non-Students. One part of the latter difference may consequently be driven by the fact that our experiment is slightly more likely to expose Politicians to

¹¹It has to be noted that all situations involving a choice at the intersection of two budget lines does not yield a CEV. For instance, in the figure A.7, if the budget line G were to cross the y -axis below the point where the budget line B crosses the x -axis, the allocations π_s^1 and π_s^2 would not constitute a GARP violation. Indeed, we would have $\pi_s^1 R_D \pi_s^2$ but not $\pi_s^2 P_D \pi_s^1$, as π_s^1 would not be affordable at prices p_s^2 .

a CEV violation (which also means that they are exposed to a slightly more powerful test).

For the sake of accuracy, we note that the negative link between Bronars (1987)'s power and the angle between intersecting budget lines is mechanical, whereas the positive link between the likelihood of a CEV and the same angle that Murphy and Banerjee (2015) observe is statistical.

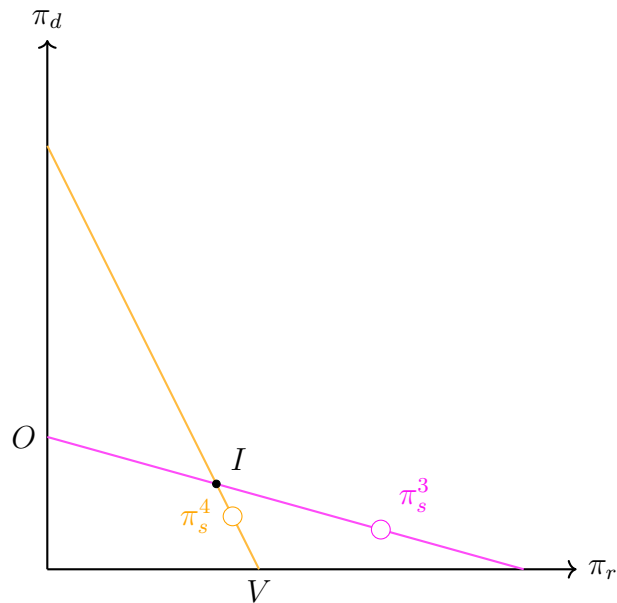
The mechanical link can be understood based on the figure A.8. There are two characteristics of the budget scenarii that influences Bronars (1987)'s power: the number of budget lines' intersections and each intersection's configuration. The former is obvious from the observation that if budget lines never intersect, then it is not possible to identify a violation of GARP, thus the experiment has zero power in detecting rational behavior. The latter is related to the parallelity of the budget lines: the probability that a random decision-maker violates GARP increases with the parallelity of budget lines. More specifically, in order to show a GARP violation in figure A.8, two events must happen:

- E^1** A choice on the budget line defined by prices \mathbf{p}_s^4 between the x -axis endpoint and dot I (i.e., strictly within the budget set defined by prices \mathbf{p}_s^3)
- E^2** A choice on the budget line defined by prices \mathbf{p}_s^3 between the y -axis endpoint and dot I (i.e., strictly within the budget set defined by prices \mathbf{p}_s^4).

Assuming independant and identically uniformly-distributed random choices along each line, the probability of each event is the norm of the segment from I to the corresponding axis divided by the norm of the entire corresponding budget line. Finally, the probability of a GARP violation is given by the product of these two probabilities ($\approx 7.15\%$ here). The higher this probability, the easier it gets to reject the null hypothesis of rational behavior, hence the more powerful the test.

Furthermore, when the slopes of the two budget lines are close (i.e., there is a small difference between the relative price), the probability of a GARP violation is bigger than when the slopes are remote. Indeed, if we fix one line (say O) and slide the other one to the left through I so that the lines are more parallel, then

Figure A.8: Possibly informative budget sets



the relative size of the segment from I to the y -axis budget line's endpoint increases (with respect to the entire V line). Thus, an increase in the parallelity of the budget lines corresponds to a power increase.

As for the link between the likelihood of a CEV violation and the angle between intersecting budget lines, it is obtained by Murphy and Banerjee (2015) from a statistical analysis. Specifically, they pool the data on GARP-violations from three existing studies (totalling 878 observations) and estimate a binomial probit model of a dichotomous variable taking value 1 if the violation is a CEV and 0 if not. Their dependant variable of interest is a proxy for the relative price differential between two intersecting budgets i and j : $2|(p_{s,d/r}^i - p_{s,d/r}^j)/(p_{s,d/r}^i + p_{s,d/r}^j)|$. They find that an increase in this proxy (hence a departure from the parallelity of budget lines) decreases the likelihood of a CEV.

However, the likelihood of a CEV is not solely determined by the position of the lines and by chance: it is also tightly connected to the subjects' actual choices (and thus to the area where these choices tend to be concentrated). As a consequence, the result may depend on the subject pool. Though the authors controlled the origin of the pool by introducing study dummies, all the studies they use are based on student

samples. If for some reason the budget lines cross near the area where students choose when they are far from being parallel while crossing away from students' choices when they are almost identical, the likelihood of a CEV will statistically increase with the relative price differential. But it may also be the case that the opposite turns true with subjects whose behavior differs from students (namely a positive relationship between the likelihood of a CEV and Bronars (1987)'s power).

We consequently replicated their analysis with our 1408 GARP violations and successfully recovered their result. Though it is quantitatively smaller (marginal effect of about -0.028 instead of -0.129 in their study), the effect of relative price differential on the likelihood of a CEV is highly significant ($p < 0.01$) and it resists the introduction of controls for the group composition and possible interaction effects. In particular, we observe that interacting the proxy variables with our three groups of interest does not yield any significant difference.

A.3 Complementary regression analyses

A.3.1 Effect of past experience in experiments

The regressions exposed in table A.3 explore the impact of Students' past experience in experiments on our rationality indices. Our "past experience" variable is directly taken from a question that is standardly added in the end-of-experiment questionnaire. Subjects could answer either that participating in our experiment was their first participation, second participation, or more. For each index, we display two models: the first model includes only the past experience variable and the second model adds available control variables.¹²

We observe a significant positive effect of experience in participating in economic experiments on rationality, no matter the index under scrutiny. Specifically, experienced subjects are both less likely to fail the GARP test and deviate less when they fail it. Interestingly, such an effect is not significant for only one prior experience in experimental decision-making. However, considering those who participated more than once in the past yields a highly-significant effect which is furthermore robust to the introduction of control variables.

Interestingly, our comparison of Politicians and Students is sensitive to removing the 32 Students (26.02%) who participated more than once in economic experiments before participating in our experiment. As shown in the table A.4, the models with only the group dummy (first column for each index) does not yield any significant difference between Politicians and Students. In other words, though the signs are consistent, we cannot reject the null that there is no difference between Politicians and Students in terms of rationality after removing experienced Students, contrary to what we observed with the whole pool of Students (table 1.4). However, once controlling for the (many) differences in observables between the two groups, we still recover that Politicians are significantly less rational than Students (second column for each index in table A.4). Yet, both the magnitude and the statistical

¹²As subjects categorized as "Economics" represent the greatest category among Students, we changed the baseline of the field of studies variable accordingly.

Table A.3: Effect of past experience in economic experiments on rationality (Students)

	AEI		HMI		VEV	
<u>Past experience</u> (Baseline: Never participated)						
Participated once	0.018 (0.037)	0.010 (0.041)	0.016 (0.021)	0.015 (0.021)	0.018 (0.036)	0.007 (0.042)
Participated twice or more	0.128*** (0.034)	0.113** (0.044)	0.083*** (0.015)	0.076*** (0.017)	0.150*** (0.045)	0.131*** (0.050)
<u>Age</u>		-0.003 (0.007)		-0.002 (0.006)		-0.003 (0.007)
<u>Gender</u> (Baseline: Male)		-0.048 (0.037)		-0.036*** (0.010)		-0.045 (0.043)
<u>Field of Studies</u> (Baseline: Economics)						
Fundamental science		-0.011 (0.043)		-0.011 (0.031)		-0.028 (0.051)
Humanities		-0.024 (0.077)		-0.003 (0.050)		-0.042 (0.097)
Other		0.009 (0.049)		0.002 (0.031)		0.005 (0.056)
<u>Higher education</u> (Baseline: No)		0.039 (0.036)		-0.002 (0.027)		0.049 (0.041)
Observations	123	123	123	123	123	123
LR chi2	14.198	53.286	29.472	37.952	13.922	41.560
Prob > χ^2	0.001	0.000	0.000	0.000	0.001	0.000

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses.

Note: AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector.

VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993). Similar results (unreported) are observed when regressing the average element in the efficiency vector.

significance¹³ of the difference are reduced as compared to the results from the full sample (table 1.4).

As a consequence, past experience in experimental decision-making appears an

¹³Reduction of the effect by 10 to 20 percentage points and statistical significance passing from strong ($p < 0.01$) to simple ($p < 0.05$) or even mild ($p < 0.1$) evidence, depending on the rationality index. We also acknowledge that the results on the efficiency vector's average are no longer consistent with the results on the minimum: the difference between Politicians and Students is no longer significant in the models without experienced Students, even when adding control variables.

Table A.4: Tobit on rationality measures (Pol v. SC ≤ 1 prior participation)

	AEI		HMI		VEV	
<u>Pol</u> (Baseline: Control)	-0.018 (0.013)	-0.030** (0.014)	-0.007 (0.005)	-0.012* (0.006)	-0.023 (0.014)	-0.032** (0.015)
<u>Gender</u> (Baseline: Male)		-0.002 (0.026)		-0.009 (0.014)		0.023 (0.030)
<u>Field of Studies</u> (Baseline: Fundamental Science)						
Economics		-0.106*** (0.040)		-0.020 (0.019)		-0.095** (0.043)
Humanities		-0.168*** (0.060)		-0.045 (0.028)		-0.186*** (0.070)
Other		-0.044 (0.031)		-0.008 (0.021)		-0.033 (0.030)
<u>Higher education</u> (Baseline: No)		0.049* (0.027)		0.022 (0.014)		0.067** (0.029)
Observations	230	230	230	230	229	229
LR chi2	2.025	18.162	2.358	8.673	2.614	16.188
Prob > χ^2	0.155	0.006	0.125	0.193	0.106	0.013
Bootstrapped p-value	0.183	0.044	0.150	0.080	0.157	0.045

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses.

Note: AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector.

VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993).

The "Bootstrapped p-value" report the p-value associated with the group variable when applying the score cluster bootstrap correction for the small number of clusters (Kline and Santos, 2012)¹⁴. The bootstrap is conducted with 4999 repetitions.

important driver of the observed differences between Politicians and Students in terms of rational behavior. In other words, it is likely that student samples are not *per se* much more consistent with the utility maximization model than other samples, but simply that they benefit from a deeper knowledge of the decision environment due to repeated participation. Indeed, participating in an economic experiment is a rather unusual experience which may require repetitions in order to grasp its functioning and be able to focus well enough. Alternatively, it may also be that participating many times entails a selection process of specific types among the population, which may correlate with choices' consistency. For instance, it may be that rational types retrieve the most of experiments because they manage to obtain precisely what they want from it without wasting their time and money

deviating from their preferences. As a consequence, they may be more likely than irrational types to value positively their participation and come back to the lab. Unfortunately, we are unable to disentangle both possible mechanisms, but the link between closeness to utility maximization and experience in economic experiments that we observe could provide an interesting avenue for future research.

A.3.2 Heterogeneity analysis of rationality within Politicians

In this section, we focus on the possible sources of the heterogeneity we observe in rational behavior among our unique sample of politicians. We are particularly interested in five possible variables which could explain Politicians' heterogeneity in rationality indices.

The most important variable is experience in policy-making, measured by the number of years spent holding local political duties. Experience in policy-making have indeed been shown to influence the quality of the politicians' pool. It has for instance been linked to dishonesty (Janezic and Gallego, 2020, Chaudhuri et al., 2020) or reciprocal behavior (Enemark et al., 2016).

The second one is the local status (Mayor, Vice-Mayor or local councillor). Even though local policies are chosen based on a vote within the local councils so that every vote counts, the level of responsibilities as well as the leadership position differ among councils' members in the municipalities. One may for instance argue that the Mayor is more exposed to the public eye and consequently more incentivized to care about his decisions than other local councillors.

The last three variables are used to question whether some standard characteristics of the underlying population matter for the choice consistency of their representatives.¹⁵ In particular, we are interested in assessing whether the size of population or its wealth have any impact. Regarding the size of the population, we look into both the (log) number of inhabitants in the town of election and whether the subject holds a term at a higher administrative layer in addition to his local term (meaning he often has to decide for a larger pool of inhabitants than his sole election town). Regarding the wealth of the population, we consider the (log) income per inhabitant or the (log) tax capacity¹⁶ per inhabitant in the town of election.

¹⁵Concerning town's characteristics, they are not self-declared but rather directly recovered from INSEE statistics based on Politicians' zip-code.

¹⁶The tax capacity measures the amount that the town could levy in local taxes had it implement average taxes at the national level. It is consequently a measure of the capacity in terms of tax resources that the town is able to extract from its territory which neutralizes the actual tax rate differential across towns.

Table A.5: Politicians' specific characteristics and rationality indices

	AEI	HMI	VEV
<u>Political experience</u> *	.105	.141	.079
<u>Population</u> (2019)*	.115	.146	.047
<u>Income per capita</u> (2019)*	.1	.095	.114
<u>Tax capacity per capita</u> (2019)*	.102	.07	.054
<u>Elected status</u>			
Mayor	.903	.926	.893
Vice-mayor	.884	.913	.88
Local councillor	.889	.916	.867
<u>Higher-layer term</u>			
No	.87	.907	.861
Yes	.902	.926	.885

Note: Entries report the Pearson's correlation coefficient between the variable and the rationality index for continuous variables (identified by *) or the average rationality index in the category for discrete variables.

AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector.

VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993).

For each variable, we look at the correlation (continuous variables) or the average per category (discrete variables) between the variable and each rationality index. The table A.5 summarizes our results. Interestingly, we observe a positive relationship between our rationality indices and all our variables: Politicians with experience in local decision-making, who are also elected at the intermunicipal level, who have a higher elected status¹⁷, or who come from a town with a more important population, income or tax capacity are slightly more rational than other Politicians. However, we uncover only two statistically meaningful relationships: experienced policy-makers and policy-makers from populous towns deviates slightly less often

¹⁷By "elected status", we mean the official position that they hold during their local mandate. Of course, this position may partly reflect another possibly more relevant characteristic which we do not observe: the degree of implication within the executive branch of the local government. Indeed, mayors and vice-mayors are *de jure* attributed some local duties, which varies from town to town, but councillors may also occasionally be attributed some duties too. Both the effect of the position *per se* and the degree of political implication are captured by our "elected status" variable.

from GARP (the Pearson correlation coefficient on the HMI is significant at the 10% level).

As a consequence, the variables we identified do not perform well in explaining the heterogeneity in rational behavior among Politicians. Yet, such exercise is necessarily limited by the sample size and would probably benefit from a nation-wide experiment involving a larger number of subjects. Finally, we note that the positive-though-insignificant relationship between holding an intermunicipal term or being elected as a Mayor and the rationality indices may partly explain the difference we observe between Lab and Field sessions. Indeed, due to the Field's recruitment process, Politicians from the Field were more often Mayors and more often seated in intermunicipal councils, which could explain the direction of the group effect (though such effect is never significant).

A.4 Complementary power analyses

A.4.1 Power of the revealed preference conditions

In this sub-section, we replicate the traditional power analysis that is presented in the majority of studies using induced budget experiments (and most notably Fisman et al., 2007, Choi et al., 2007). Such power analysis is based on Bronars (1987)'s approach and consists in running a series of Monte-Carlo experiments with uniform-random subjects. Specifically, we simulate 25,000 subjects entering the lab exactly the same way actual subjects did: each faces 30 randomly-generated budget lines with the same restrictions on budget lines' endpoints and "chooses" an allocation at random along each line. For each simulated subject, we compute the various indices presented in the last sections and then compare the results from these simulated subjects with the results from our actual subjects. The results are depicted in figure A.9.

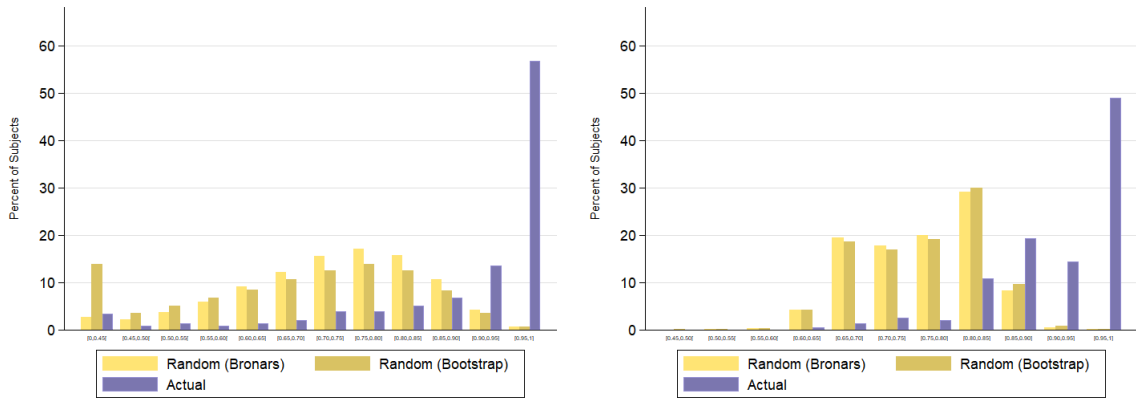
We also run simulations based on the bootstrap approach proposed by Andreoni and Miller (2002) and Andreoni et al. (2013). Under this approach, instead of picking allocations at random with equal probabilities, each simulated subject is attributed at random for each of her "choices" a budget share that was chosen by one of the actual subjects. Such a procedure thus accounts for the structure imposed by actual choices on the power of the test and not only on the internal structure entailed by the budget lines. Indeed, the uniform benchmark may misestimate power if the possible violations implied by the choices' pattern are actually impossible to detect, even though some other types of violations can be detected. Furthermore, considering our groups may behave differently, we split the 25,000 simulated subjects into sub-groups in a way that preserves the proportions observed in our actual sample and then randomize budget shares within each corresponding group.¹⁸

¹⁸Our procedure is actually closer to Dean and Martin (2016)'s bootstrap implementation than to Andreoni and Miller (2002) and Andreoni et al. (2013)'s proposal. The difference relies on bootstrapping only on identical budget environments in the latter case (which is not possible in our experiment since there are no subjects confronted to the exact same budget line) against bootstrapping on all budget environments in the former case. Dean and Martin (2016)'s approach basically amounts to projecting observed choice on unobserved budget sets and consequently implies some form of consistency between observed and unobserved choices.

APPENDIX A

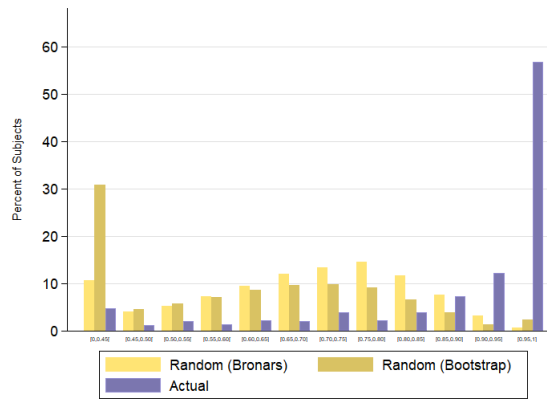
In all cases, actual subjects are much closer to GARP than simulated subjects. For instance, there are about 70% of participants who have an AEI above 0.9, against about 5% (resp. 4%) of simulated subjects using Bronars' (resp. bootstrap) method.

Figure A.9: Statistical power analysis



(a) AEI

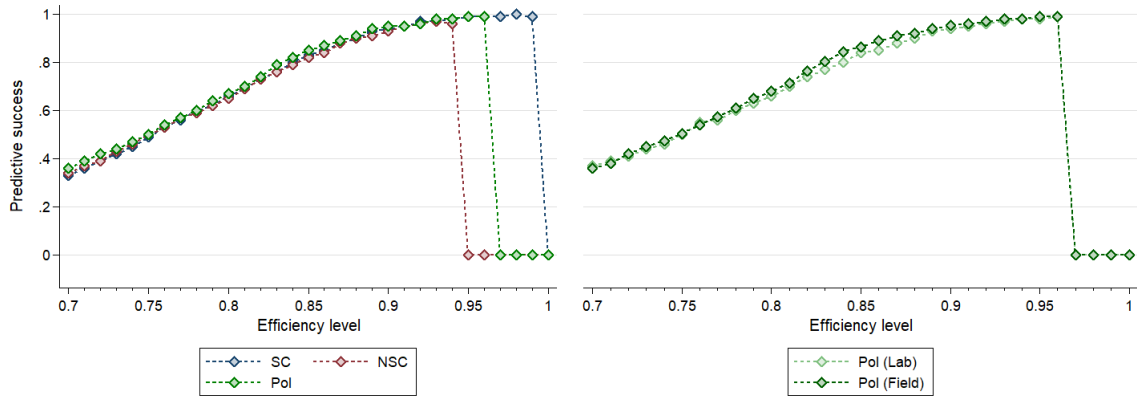
(b) HMI



(c) VEV

The figure A.10 is the counterpart of the figure 1.4 which substitutes the average predictive power with the median predictive power: for each efficiency level, the graph plots the median predictive power within the corresponding group.

Figure A.10: Median predictive success for different efficiency levels across groups



(a) Main groups

(b) Politicians sub-groups

The table A.6 exposes the p-values from statistical tests comparing the distribution of the Selten scores between our groups. Globally, we observe the same results as the one exposed for the rationality indices, except the difference between Politicians and Students are less striking. In other words, once we take into account the possible difference between the groups in the stringency of the revealed preference tests, our results remain globally unchanged.

Table A.6: p -values from statistical tests for treatment effects on Selten scores

	Pol v. NSC			Pol v. SC			Lab v. Field		
	AEI	HMI	VEV	AEI	HMI	VEV	AEI	HMI	VEV
t-test	.129	.581	.046	.037	.072	.025	.483	.968	.149
FP test	.338	.604	.151	.468	.262	.328	.725	.914	.333
ES test	.509	.971	.415	.233	.084	.02	.872	.68	.348
Observations	230	230	229	262	262	261	139	139	138

Note: AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector. t-test: two-samples Student test. FP: Fligner-Policello. ES: Epps-Singleton. VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993). Similar results (unreported) are observed when comparing the average element in the efficiency vector.

A.4.2 Power of the statistical tests

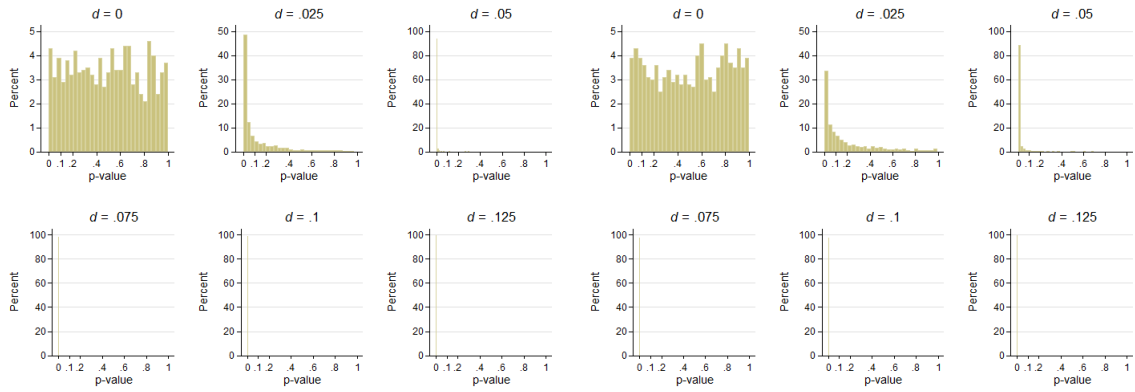
This sub-section describes the simulation procedure mentioned in section 4 to assess the power of the statistical tests we implement to detect a difference across experimental groups when no difference is actually detected. Such procedure borrows from Müller (2019) and is extended to the tests we use. In any case, given a behavioral measure computed for one baseline group (b) and one "treated" group (t) with sample sizes n_k , empirical means μ_k and empirical standard deviations σ_k ($k \in b, t$), we repeat a hundred times the following steps for each "treatment effect" $d \in \{0, 0.025, 0.05, 0.075, 0.1, 0.125\}$:

1. Draw n_b iid observations from a normal distribution with mean $\hat{\mu}_b$ and standard deviation $\hat{\sigma}_b$.
2. Draw n_t iid observations from a normal distribution with mean $\hat{\mu}_t \pm d$ and standard deviation $\hat{\sigma}_t$ (the sign of d depends on the direction of the empirical difference between the groups).
3. Replace draws lying outside of the behavioral measure's boundaries with the corresponding boundary (e.g., the AEI is necessarily below 1 so that any draw strictly above 1 should be restricted to 1).
4. Store the p-values from applying the statistical tests to detect differences between the two simulated distributions.

We implement this procedure for all our rationality indices. We restrict our attention to situations where all implemented tests do not detect a difference between two groups (in particular Politicians v. Non-Students or Politicians from the Lab v. Politicians from the Field). Distributions of p-values from Fligner-Policello tests on the differences between our groups are depicted in figures A.11 and A.12 (along with figure 1.5 in section 4). The results for all behavioral measures are summarized in table A.7.

APPENDIX A

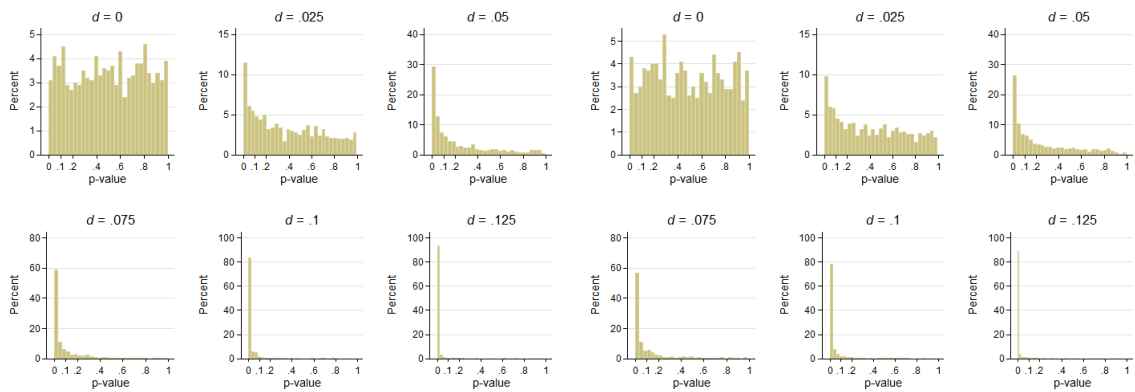
Figure A.11: P-values from simulations of a hundred Fligner-Policello tests on differences in HMI across groups



(a) Pol v. NSC

(b) Lab v. Field

Figure A.12: P-values from simulations of a hundred Fligner-Policello tests on differences in VEV across groups



(a) Pol v. NSC

(b) Lab v. Field

Table A.7: Summary statistics for statistical tests based on simulated data

(a) Pol v. NSC

		$d = 0$		$d = 0.025$		$d = 0.05$		$d = 0.075$		$d = 0.1$		$d = 0.125$	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
AEI	t-test	0.44	0.30	0.27	0.29	0.09	0.17	0.02	0.07	0.00	0.02	0.00	0.00
	FP test	0.50	0.29	0.39	0.30	0.18	0.24	0.06	0.14	0.01	0.06	0.00	0.01
	ES test	0.30	0.27	0.25	0.24	0.15	0.19	0.07	0.12	0.02	0.04	0.00	0.02
HMI	t-test	0.48	0.30	0.10	0.18	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	FP test	0.49	0.29	0.13	0.20	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	ES test	0.45	0.28	0.19	0.21	0.01	0.05	0.00	0.00	0.00	0.00	0.00	0.00
VEV	t-test	0.42	0.31	0.24	0.28	0.10	0.18	0.03	0.09	0.01	0.03	0.00	0.01
	FP test	0.50	0.29	0.39	0.30	0.22	0.26	0.09	0.17	0.02	0.07	0.01	0.04
	ES test	0.21	0.23	0.18	0.21	0.13	0.18	0.06	0.11	0.02	0.06	0.01	0.03

(b) Lab v. Field

		$d = 0$		$d = 0.025$		$d = 0.05$		$d = 0.075$		$d = 0.1$		$d = 0.125$	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
AEI	t-test	0.47	0.29	0.35	0.31	0.16	0.23	0.06	0.11	0.01	0.06	0.00	0.02
	FP test	0.49	0.30	0.40	0.30	0.23	0.26	0.09	0.16	0.03	0.08	0.01	0.04
	ES test	0.40	0.28	0.34	0.28	0.23	0.23	0.13	0.18	0.05	0.11	0.02	0.06
HMI	t-test	0.51	0.30	0.19	0.24	0.02	0.05	0.00	0.00	0.00	0.00	0.00	0.00
	FP test	0.51	0.30	0.20	0.25	0.01	0.06	0.00	0.00	0.00	0.00	0.00	0.00
	ES test	0.48	0.28	0.27	0.25	0.04	0.09	0.00	0.01	0.00	0.00	0.00	0.00
VEV	t-test	0.44	0.31	0.27	0.28	0.13	0.19	0.04	0.11	0.01	0.03	0.00	0.01
	FP test	0.49	0.29	0.41	0.30	0.25	0.27	0.11	0.19	0.03	0.09	0.01	0.03
	ES test	0.24	0.24	0.20	0.21	0.15	0.19	0.08	0.13	0.03	0.08	0.01	0.04

Note: AEI: Afriat's Efficiency Index. HMI: Houtman-Maks Index. VEV: Varian's Efficiency Vector. t-test: two-samples Student test. FP: Fligner-Policello. ES: Epps-Singleton. VEV here stands for the minimum in the efficiency vector, following suggestion by Varian (1993).

References

- Afriat, S. N. (1967). The construction of utility functions from expenditure data. International Economic Review, 8(1):67–77.
- Afriat, S. N. (1972). Efficiency estimation of production functions. International Economic Review, pages 568–598.
- Afriat, S. N. (1973). On a system of inequalities in demand analysis: an extension of the classical method. International Economic Review, pages 460–472.
- Afriat, S. N. (1987). Logic of choice and economic theory. Oxford: Clarendon Press.
- Alcantud, J. C. R., Matos, D. L., and Palmero, C. R. (2010). Goodness-of-fit in optimizing a consumer model. Mathematical and Computer Modelling, 52(7-8):1088–1094.
- Andreoni, J., Gillen, B. J., and Harbaugh, W. T. (2013). The power of revealed preference tests: Ex-post evaluation of experimental design. Unpublished manuscript.
- Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. Econometrica, 70(2):737–753.
- Bronars, S. G. (1987). The power of nonparametric tests of preference maximization. Econometrica: Journal of the Econometric Society, pages 693–698.
- Chaudhuri, A., Iversen, V., Jensenius, F. R., and Maitra, P. (2020). Selecting the best of us? politician quality in village councils in west bengal, india. Technical report, CESifo Working Paper.
- Cherchye, L., De Rock, B., Sabbe, J., and Vermeulen, F. (2008). Nonparametric tests of collectively rational consumption behavior: an integer programming procedure. Journal of Econometrics, 147(2):258–265.
- Cherchye, L., Demuynck, T., De Rock, B., and Hjertstrand, P. (2015). Revealed preference tests for weak separability: an integer programming approach. Journal of Econometrics, 186(1):129–141.
- Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007). Consistency and heterogeneity of individual behavior under uncertainty. American Economic Review, 97(5):1921–1938.

- Cox, J. C. (1997). On testing the utility hypothesis. The Economic Journal, 107(443):1054–1078.
- Cox, J. C. (2004). How to identify trust and reciprocity. Games and Economic Behavior, 46(2):260–281.
- Dean, M. and Martin, D. (2016). Measuring rationality with the minimum cost of revealed preference violations. Review of Economics and Statistics, 98(3):524–534.
- Di Bartolomeo, G. and Papa, S. (2016). Trust and reciprocity: extensions and robustness of triadic design. Experimental Economics, 19(1):100–115.
- Enemark, D., Gibson, C. C., McCubbins, M. D., and Seim, B. (2016). Effect of holding office on the behavior of politicians. Proceedings of the National Academy of Sciences, 113(48):13690–13695.
- Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. American Economic Review, 94(4):857–869.
- Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. American Economic Review, 97(5):1858–1876.
- Gross, J. and Kaiser, D. (1996). Two simple algorithms for generating a subset of data consistent with warp and other binary relations. Journal of Business & Economic Statistics, 14(2):251–255.
- Halevy, Y., Persitz, D., and Zrill, L. (2018). Parametric recoverability of preferences. Journal of Political Economy, 126(4):1558–1593.
- Heufer, J. and Hjertstrand, P. (2015). Consistent subsets: Computationally feasible methods to compute the houtman–maks-index. Economics Letters, 128:87–89.
- Heufer, J. and Hjertstrand, P. (2019). Homothetic efficiency: Theory and applications. Journal of Business & Economic Statistics, 37(2):235–247.
- Houtman, M. and Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. Kwantitatieve methoden, 19(1):89–104.

- Janezic, K. A. and Gallego, A. (2020). Eliciting preferences for truth-telling in a survey of politicians. Proceedings of the National Academy of Sciences, 117(36):22002–22008.
- Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. Journal of Econometric Methods, 1(1):23–41.
- Müller, D. (2019). The anatomy of distributional preferences with group identity. Journal of Economic Behavior & Organization, 166:785–807.
- Murphy, J. H. and Banerjee, S. (2015). A caveat for the application of the critical cost efficiency index in induced budget experiments. Experimental Economics, 18(3):356–365.
- Smeulders, B., Spijksma, F. C., Cherchye, L., and De Rock, B. (2014). Goodness-of-fit measures for revealed preference tests: Complexity results and algorithms. ACM Transactions on Economics and Computation (TEAC), 2(1):1–16.
- Varian, H. R. (1982). The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society, pages 945–973.
- Varian, H. R. (1984). The nonparametric approach to production analysis. Econometrica: Journal of the Econometric Society, pages 579–597.
- Varian, H. R. (1990). Goodness-of-fit in optimizing models. Journal of Econometrics, 46(1-2):125–140.
- Varian, H. R. (1993). Goodness-of-fit for revealed preference tests. University of michigan ann arbor, Department of Economics.

Chapter 2

The Distributional Preferences of Politicians: Experimental Evidence¹

Contents

1	Introduction	132
2	Literature-based behavioral conjectures	140
3	Experimental design	153
4	Distributional preferences analysis	159
5	Results	166
6	Discussion	189
7	Concluding remarks	194
	References	198

¹This chapter is an extended version of a working paper co-authored with Laurent Denant-Boèmont & Matthieu Leprince.

”Political writers have established it as a maxim, that, in contriving any system of government, and fixing the several checks and controuls of the constitution, every man ought to be supposed a knave, and to have no other end, in all his actions, than private interest. By this interest we must govern him, and, by means of it, make him, notwithstanding his insatiable avarice and ambition, co-operate to public good. Without this, say they, we shall in vain boast of the advantages of any constitution, and shall find, in the end, that we have no security for our liberties or possessions, except the good-will of our rulers; that is, we shall have no security at all.”

David Hume, Essays Moral, Political, Literary, 1777

1 Introduction

Decisions made on a regular basis by politicians have a tremendous impact on the daily lives of their constituents. In particular, most decisions from taxation choices to budget programming and public service provision boil down to allocating economic resources across groups of individuals, hence a major redistributive power with substantial impacts on the aggregate welfare of the nation. When citizens have perfect control over what politicians do or when proper institutions and incentives prevent deviations from the population’s demands, the identity of politicians should not matter to policy-making. Yet, as control is limited and institutions often fail, politicians’ qualities ultimately matter (Braendle, 2016). Indeed, a growing number of studies emphasizes the importance of politicians’ individual characteristics for policy-making (e.g., the literature review in Kuliomina, 2021). Understanding who are politicians and what drives their behavior is therefore a key element in the understanding of political decisions. However, the majority of studies from this strand of literature focuses on physical and psychological traits rather than on behavioral motives and personal motivations, which is precisely the focus of the present study.

In economics and political science, the usual modelization of politicians’ behavior relies on the assumption that the decision-maker is a ”knave” who will act only out of his best interest. Political institutions (and foremost democratic elections) are then conceptualized as a curbing mechanism preventing self-interested decision-makers to

earn a personal profit out of their position (Downs, 1957).² When the institutions fail, politicians are thus able to hijack the public good and fulfil their own agenda. Yet, this agenda is generally assumed to include only extrinsic motivations, so that the remedies to prevent politicians from behaving badly do not account for and even offset existing intrinsic motivations.³

At the same time, there are several reasons why intrinsic motivations are likely essential in the understanding of politicians' behavior.

First, in the absence of the possibility to exert a perfect external control on politicians' behavior (through institutions and incentives) or because of the impossibility to determine *ex-ante* the desired policy outcome in many situations, voters may select politicians based on their personal characters rather than on their political platforms (Lomasky, 2008, Dal Bó and Finan, 2018). Consequently, the intrinsic values or "principles" of candidates may be important features of an election, because they partly determine what voters perceive as "good types" (Fearon, 1999, Gulzar and Khan, 2021).⁴ In addition, scrutinizing politicians moral values is of particular interest given their possible retroactive effect on moral values in the entire population. Indeed, politicians are not only decision-makers to which the execution of the public will have been delegated: they also hold characteristics of leaders in the political arena. Consequently, they may act as "role models" (Gächter et al., 2012, Gächter and Renner, 2018) or even moral "entrepreneurs" (Mansbridge, 2003, Schnellenbach and Schubert, 2015) and influence citizens' view about what constitutes moral behavior.

²Throughout this text, we will equate "self-interest" with the maximization of one's own private benefit. Surely, "self-interest" for politicians may not directly reduce to material wealth (e.g., seeking reelection, maximizing budgets, optimizing fame, etc.). However, the wealth-maximizing decision-maker remains the standard conception upon which both political economy and public choice theory are built (Kirchgässner, 2014).

³Following Frey (1992), who quotes seminal psychological definition by Deci (1971), intrinsic motivations can be defined as follows: "one is said to be intrinsically motivated to perform an activity when one receives no apparent reward except the activity itself". In this case, behavior is based on the moral and ethical considerations which form part of people's preferences". Rewards other than the activity itself (e.g., the monetary benefits derived from the activity) thus correspond to "extrinsic" motivations.

⁴In political representation theory, the corresponding concept is "gyroscopic representation": "In this model of representation, voters select representatives who can be expected to act in ways the voter approves without external incentives. The representatives act like gyroscopes, rotating on their own axes, maintaining a certain direction, pursuing certain built-in (although not fully immutable) goals." (Mansbridge, 2003).

Second, internal values also likely matter for the self-selection into politics. Indeed, the decision to hold a public office has been shown to imply a very specific kind of motivations, namely "public service motivations", which is commonly defined as "an individual's predisposition to respond to motives grounded primarily or uniquely in public institutions and organizations" (Perry and Wise, 1990). Public service motivations may even serve as a justification for government provision of public services, as opposed to public service delegation (Francois, 2000). Although the bulk of selection models (e.g., Besley and Ghatak, 2005, Delfgaauw and Dur, 2008) and empirical investigations (e.g., Lewis and Frank, 2002, Prendergast, 2007, Georgellis et al., 2011, Dal Bó et al., 2013) focus on civil servants, applying this concept to politicians is straightforward. For instance, Besley (2005) conceptualizes the selection of political candidates based on a tension between rent-extraction and public service motivations. The rationale is similar to the one used to explain why public organizations may attract highly-capable workers even when such workers should shy away from relatively poorly-paid jobs as compared to the private sector: there exists non-extrinsic rewards to participate in the public good (e.g., commitment to serving the public interest, compassion, self-sacrifice, desire to be useful to society, civic duty achievement, etc.). Such a statement joins seminal contributions from sociology which describes politicians as vocational agents fulfilling a call to do things right and actively contribute to society (Weber, 1918). Empirically, several studies emphasize the importance of such motives for politicians. In a field experiment, Broockman (2013) observes that african-american politicians are more intrinsically motivated to advance the welfare of african-american citizens than their white counterparts. Ritz (2015) finds that public service motivations matter for various political activities (volunteering, seeking reelection, executive work, etc.), though with differing intensity depending on the activity (also see Pedersen, 2014). Relying on an innovative design that manipulates the type of incentives (personal returns or pro-social motivations) to run for political candidacy in a field experiment in Pakistan, Gulzar and Khan (2021) find pro-social motivations to increase the likelihood to both run and get elected, as well as to improve policy alignment with citizens' preferences.

Third, it is now a well-established result from the behavioral and experimental economics literature that individuals do not solely maximize their own private well-being but also incorporate others' well-being (Camerer, 2011, Cooper and Kagel, 2016). In particular, when asked to divide an amount of money, many people are

willing to pay to deviate from their own gains and favor alternative motives for distribution, or so-called "distributional preferences".⁵ To the extent that politics is primarily concerned about distributing some economic resources across competing uses, hence competing groups of people, these results quite naturally extend to political decisions. Indeed, the idea that politicians deviate from voters' preferences and implement their own preferences still stands even when we do not restrict motivations to the sole maximization of personal well-being. Besides, politicians' moral values likely matter to policy-making precisely because citizens' preferences are not necessarily available to politicians at any time. Indeed, politicians only have incomplete information about voters' preferred policies (Lindbeck and Weibull, 1987, Coughlin, 1992): voters do not necessarily know what they want (and it is indeed partly the purpose of electing representatives, see Mansbridge, 2003), and even when they do know they do not necessarily convey this information to policy-makers. Ultimately, politicians should pursue the public good, understood as what the citizens want, but what constitutes such public good is uncertain. As a consequence, politicians shall act out of what they think is the public good, which will depend on their personal beliefs about what is "just" or "fair".

Our major contribution is to focus on politicians' intrinsic motivations, and in particular on distributional concerns.⁶ We indeed assume (and verify) that intrinsic motivations in general, and distributional preferences in particular, are important drivers of politicians' behavior that are also likely to differ from those of the remainder of the population. To do so, we rely on the experimental economics methodology which allows us to finely disentangle the motives behind distribution choices at the individual level. Specifically, we evaluate how politicians weigh themselves as opposed to other distributional motives (namely favoring others' well-being, favoring reducing inequalities, or favoring maximizing the social surplus). Our study thus

⁵"Distributional preferences" refers to the sub-class of other-regarding preferences which restricts the attention to preferences over the allocation of resources that affect other persons besides the decision-maker, absent any type of reciprocity-based motivations (see Kerschbamer, 2015, for a brief typology of other-regarding preferences and an overview of existing distributional preferences models).

⁶In a purely distributional context without any strategic interactions, deviations from self-interest in the form of "distributional preferences" (hence positive valuation of non-extrinsic rewards) can be interpreted as intrinsic motives if one abides by Deci (1971)'s definition (see footnote 3). The link between intrinsic and pro-social motivations is also discussed in Besley and Ghatak (2018).

provides an empirical assessment of the relative importance of private gains versus intrinsic motivations among actual politicians in a controlled environment.

Throughout this chapter, we will focus on three types of distributional motives/intrinsic motivations: Altruism (i.e., the weight attached to others' well-being), Equality (i.e., the sensitivity to reducing inequalities) and Efficiency (i.e., the sensitivity to increasing the social surplus). Altruism is the direct counterpart of Selfishness (i.e., the weight attached to personal well-being), so that investigating how politicians tradeoff both motives appears a rather intuitive inquiry. The focus on the Equality/Efficiency tradeoff is derived from the empirical relevance of such a tradeoff in many policy decisions. Indeed, pondering Equality against Efficiency has been described as the "big tradeoff" faced by governments (Okun, 1975): governments cannot implement a fully efficient redistributive policy (i.e., Equality) without reducing the aggregate income (i.e., Efficiency) because part of the funds allocated to redistribution are likely to "leak" (e.g., administrative costs). As a consequence, politicians, more than any ordinary citizen, are daily confronted to pondering an equitable distribution of resources against the maximization of these resources. Ultimately, how they solve this dilemma has a much deeper impact on people's welfare than any individual decision made by the same people. Furthermore, an important increasing body of experimental studies emphasize that both Equality and Efficiency matter to subjects in distribution games (among many others: Fehr and Schmidt, 1999, Charness and Rabin, 2002, Andreoni and Miller, 2002, Konow, 2003, Engelmann and Strobel, 2004). Since politicians' daily activity mostly involve distributing a given amount of budget across several uses, investigating Equality and Efficiency among politicians appears worthwhile.⁷

To the best of our knowledge, our study is the very first one which both investigates distributional concerns among politicians (and particularly the trade-off between Equality and Efficiency) and accounts for individual heterogeneity. There may be several reasons explaining the scarcity of research on politicians' motivations in economics.

⁷We acknowledge that pitting Equality against Efficiency is debatable: policy-makers may be able to achieve both in some situations (Boadway and Keen, 2000). However, to the extent that a tradeoff likely exists, investigating it appears valuable. We provide an argument in favor of examining the Equality/Efficiency tradeoff for politicians with greater details in the next section.

First, as a methodological issue, relying on observational data to infer politicians' preferences is challenging. Indeed, actual political decisions result from a collective process which involves a wide variety of agents (politicians, voters, lobbies, public servants, regulators, etc.) and some aggregation rule (basically a voting rule). Such a feature causes a serious identification issue. In addition, policies are generally multi-dimensional. On the one hand, such a feature complexifies the task of linking one individual's intrinsic motivation to a specific outcome. On the other hand, it also entails that understanding the motives behind actual political decisions requires adopting a global perspective. However, only a fraction of political decisions are generally observed due to constraints on data availability (e.g., only a specific kind of expenditures or only some taxation choices). Finally, the complexity of observational data does not generally allow to disentangle the motives behind the decisions of a single individual, even when observed. Strategic behavior may indeed bias the analysis: politicians may target specific groups from the population in order to obtain an electoral advantage (e.g., invest resources in swing voters or core voters, indulge in pork-barrel policies, bargain to reduce ideological oppositions, or else ; see the review on distributive politics by Golden and Min, 2013). Such a behavior is then likely to be confounded with other distributional concerns. The very complexity of the decision environment may even reinforce self-serving behavior on the part of politicians, because it eases self-deception and motivated reasoning (Zamir and Sulitzeanu-Kenan, 2018).⁸ As a consequence, politicians may not be more selfish than the remainder of the population but still act as if they were because their decision environment enhances self-serving behavior. And even when observed behavior is not selfish but corresponds to the application of an allocation rule, there simply exists too many such rules and combination of rules (equality of public inputs or outcomes, equity or efficiency considerations of various forms, discriminations or favoritism, etc., see Shoup, 1989) to allow for any truthfully-meaningful analysis. Looking directly at the result of politicians' decisions (policies) may thus be uninformative as to the real preferences of politicians, and perhaps even misleading. One example of the difficulty to derive robust conclusions as to the underlying mo-

⁸Self-deception refers to the fact that people rationalize their immoral behavior in a way that preserves their seemingly moral character (e.g., blaming the circumstances or others, emphasizing a worst-case scenario, justifying based on alternative morale values, etc.). Motivated reasoning refers to the fact that people process information in a way that serves their interest, hence not necessarily in a way that maximizes predictive accuracy.

tives of politicians is congruence with the median voter. Indeed, it could be that self-interested election-seeking politicians cater to voters' preferences. But it could also be that public-service-motivated politicians genuinely believe that implementing what voters want is the right thing to do. Both motives achieve the same result but are conceptually opposed and would yield very different and possibly contradictory recommendations as to how to achieve congruence when not observed.

Second, as a logistical issue, directly asking politicians is often very demanding. Indeed, establishing first contact with politicians is difficult due to multiple barriers.⁹ Besides, their opportunity costs to participate in a research study are arguably high given their tight schedules. Reputation fallouts may also deter them from participating (in particular if they know that their personal motivations will be scrutinized). As emphasized by Vis and Stolwijk (2020), politicians may even delegate to their administrative staff their responses to online surveys, hence a major measurement bias. Due to these difficulties, existing experimental studies on political elites often have a wide definition of their subject pools (e.g., politicians from all governmental tiers, administrative or military staffs, doctor in economics, individuals with education and/or income above a certain threshold, etc.) which ultimately blend very heterogeneous elite profiles within the same study (Kertzer and Renshon, 2022). Such a practice may subsequently blur the correct interpretation of existing findings.

Third, as an ontological issue, the way politicians as individuals matter for policy-making in economics has essentially been conceptualized and defended by the Public Choice school. Public choice scholars depict politicians as rational utility maximizers against the "benevolent planner" from the previously-dominant welfarist approach. Although what constitutes such utility may theoretically encompass a wide array of motivations, and despite some early calls to incorporate behavioral insights into the analysis of politicians' behavior (e.g., Simon, 1995, Kliemt, 2005, Wilson, 2011), most of the existing literature still builds upon an egoistic and materialistic premise. Contrary to behavioral approaches, the public choice's modelization provides an uni-

⁹These include (non-comprehensive selection): preparing a professional-looking yet simple introductory message ; obtaining ideally the support of one political organization (associations, intermunicipal/departamental/regional structures, etc.) ; mobilizing personal networks to facilitate contact ; phoning and/or mailing the secretariat of politicians' institutions and then possibly their personal secretariat too ; following up several/many times to make sure the message has not been lost ; etc. For other examples, see Vis and Stolwijk (2020).

ifying theory of political behavior with stable testable predictions (Wallerstein, 2004). Given the complexity of the environment surrounding political decision-making, this feature likely explains the public choice school's attractiveness and its rapid institutionalization, which may in turn have shelved a questioning about politicians' actual motivations and behavior in economics (Schnellenbach and Schubert, 2015).

We address these limitations and fill in the existing gap in the literature by recruiting actual politicians to participate in a distribution experiment. The experimental economics methodology indeed supplies a valuable framework to investigate politicians' intrinsic motivations: all choices happen within a tightly-controlled decision environment which involves an explicit incentive scheme. Intrinsic motivations are therefore gauged against the most classical extrinsic motivation (monetary gains) in a straightforward way.

Specifically, we use a modified version of the dictator game (Andreoni and Miller, 2002, Fisman et al., 2007) to accurately disentangle two independent components of distributional preferences: how subjects evaluate their own versus other's well-being (Selfishness/Altruism tradeoff) and how subjects trade off Equality (reducing income differences) against Efficiency (increasing total income)¹⁰ on the other hand. Our empirical strategy relies on the conjunction of revealed preference techniques and the estimation of a structural utility function to obtain individual parameters identifying both tradeoffs. Based on such estimated parameters, we compare the distributional preferences of our sample of 139 politicians with the preferences from two control groups: a convenience student sample (123 subjects) and a sample from the general population with demographics comparable to politicians' ones (91 subjects). In addition, we replicate our experiment in the field to investigate whether politicians coming in the lab (67 subjects) and politicians recruited in the field (72 subjects) behave differently.

¹⁰Though standard in the experimental literature on distributional preferences (e.g., Charness and Rabin, 2002, Engelmann and Strobel, 2004), such a definition of "efficiency" does not necessarily overlap with the standard paretian concept. Indeed, a frequent situation in our game is one where a subject increases the total payoff (hence efficiency in our sense) by simultaneously increasing the payoff of one subject and decreasing the payoff of another subject. Such a behavior does not consequently lead to a Pareto-efficient allocation. Though a possibly fruitful line of research, investigating Pareto-efficiency as a behavioral motive is not the present study's focus. For an example of distribution experiment that inspects sensitivity to Pareto-efficiency, we refer the interested reader to Cetre et al. (2019) and references therein.

We find stark differences in observed behavior between politicians and both our control groups. Specifically, politicians show an overwhelming tendency to favor the 50/50 split: more than 55% of our sample roughly equalizes the payoffs, which is 25 percentage points more than non-politicians and 35% more than students. Such a general pattern is driven by both a higher weight placed on others and for a higher concern for equality considerations. In particular, we observe a lower concentration of near-selfish subjects and an increase in fair-minded subjects (equal weights) among politicians as compared to both control groups. Overall, politicians tend to be biased towards themselves, but the proportion of selfish types is approximately the same as the proportion of other types (fair-minded and altruistic). Besides, politicians are predominantly equality-oriented (two-thirds of our sample), whereas non-politicians are globally balanced with respect to the Equality/Efficiency dilemma and students favor efficiency. We also find little differences between our lab and our field experiments, which increases the external validity of our experimental measures. Finally, our results resist the introduction of standard observables and cannot be explained either by differences in consistency with the underlying utility model based on testing the proximity with revealed preference axioms or by differences in prediction errors during the structural estimation process. These findings complement the existing literature on intrinsic motivations on the part of politicians by underlining their prevalence in a controlled incentivized laboratory experiment.

The remainder of the chapter proceeds as follow. Section 2 derives some behavioral conjectures from the existing literature. Section 3 depicts the experiment in details. Section 4 exposes our distributional preference analysis at the individual-level. Section 5 exposes the main results based on between-groups comparisons. Section 6 proposes a discussion on the results. Finally, section 7 summarizes and concludes.

2 Literature-based behavioral conjectures

In this section, we review the literature that allows to formulate hypotheses on how politicians would trade-off Selfishness against Altruism on the one hand, and Equality against Efficiency on the other hand. Building on such review, we propose

behavioral conjectures as to how subjects would behave in our experiment and to what extent our experimental groups would differ. We consider the literature on each tradeoff separately and then compile all insights into a bunch of conjectures.

2.1 Selfishness/Altruism tradeoff

Finding examples of seemingly self-interested decisions by politicians in both the academic and the public literature does not require much effort. Indeed, several studies emphasize that politicians may be able to extract substantial private benefits from office holding, such as increases in personal wealth (Eggers and Hainmueller, 2009, Fisman et al., 2014, Truex, 2014), revolving-door opportunities (Blanes i Vidal et al., 2012, McCrain, 2018), employment opportunities for relatives (Fafchamps and Labonne, 2017, Gagliarducci and Manacorda, 2020), or sustained power capture (Dal Bó et al., 2009, Folke et al., 2016). Some of these benefits may even involve embezzlement of public funds or other forms of corruption (Reinikka and Svensson, 2004, McMillan and Zoido, 2004, Ferraz and Finan, 2011). Even in developed countries there exists many examples of public scandals involving politicians misbehaving.¹¹ Furthermore, a large literature emphasizes that political decisions cater to the interests of office holders or to those of influent lobbies to the detriment of the general public (e.g., Jacobs and Page, 2005, de Figueiredo Jr and Edwards, 2007, Gilens and Page, 2014, Schnellenbach and Schubert, 2015).

However, as previously emphasized, these findings do not ultimately indicate that politicians act selfishly, nor that the most selfish types among the population self-select into political office. Indeed, it may simply be that office holders are at worst opportunistic and favor their own interest when it is possible, just like non-office holders would do in a similar situation. In other words, politicians may simply have more opportunities to advance their private well-being. At the same time, there also exists rationales for why politicians may be more concerned about their personal well-being than non-politicians, absent any contextual effects.

¹¹In France, one of the most famous examples in recent years is certainly the Balkany story. The Balkany spouses were local politicians convicted of multiple tax frauds and misappropriation of public funds. After some prison time, they were placed under electronic supervision at home due to health issues, but ultimately returned to prison after multiple violations of their allowed perimeter of movement.

First, politicians are primarily electoral competitors that managed to outplay their contenders. Yet, winning a contest provides a sense of entitlement which increases the likelihood that winners behave selfishly (Schurr and Ritov, 2016). In a modified dictator game in India varying the degree of decisions' publicness, Banerjee et al. (2020) find politicians to give less when the decisions are disclosed. The authors argue that such a stylized fact is indeed compatible with an entitlement interpretation: politicians, because they have been elected to an influential position within society, feel entitled to more money when they can be identified by other people.

Second, existing research in personality find that successful political leaders sometimes show political traits that associate with selfish behavior, such as disagreeableness (Best, 2011, Joly et al., 2019).¹² However, such a finding is not consensual in this literature and is likely dependant on the political context. For instance, Caprara et al. (2010) find Italian politicians to score higher on "Energy, Agreeableness and Social Desirability" than the remainder of the population, Hanania (2017) observes "Extraversion, Agreeableness, Emotional Stability and Conscientiousness" among American politicians, and Scott and Medeiros (2020) emphasize "Extraversion, Openness and Emotional Stability" as key drivers of political success in Canada.

Third, some recent studies also directly observe policy-makers' behavior at the individual level using economic experiments. Enemark et al. (2016) do not find any statistical difference in a standard dictator game between close winners and close losers of a local election in Zambia (and though they do not compare politicians with "ordinary" citizens, the distribution of the share of endowment sent in the game is fairly close to what is observed in the literature with non-politician samples). Likewise, Banerjee et al. (2020) do not document a difference between politicians and non-politicians when decisions are concealed. Using elite samples (graduates from Yale as well as sub-samples from the general American population), Fisman et al. (2015b) find American elites to be less fair-minded than the remainder of the American population in basically the same experiment as ours. In addition, there

¹²For an analysis on the link between personality traits and distributive behavior, see Ben-Ner et al. (2008). In particular, they find that Agreeableness is positively correlated with sending money in a standard dictator game, whereas the positive relationship with Extraversion (i.e., the propensity to be interested in something external to self) is sensitive to whether the decision is hypothetical or not.

also exists evidence that surviving candidates over several political cycles engage in morally-reprehensible behavior, such as cheating and lying (see Janezic and Gallego, 2020, Chaudhuri et al., 2020).

Along these lines of thoughts, one should therefore expect politicians to be at least as prone to behaving selfishly as the remainder of the population, if not more (*selfishness hypothesis*).

Alternatively, other mechanisms that have been highlighted mostly in the literature outside of economics suggest that politicians could display a greater concern for others than the remainder of the population (*altruism hypothesis*).

The previously-mentioned literature on public service motivation suggests that politicians are likely to hold pro-social concerns.¹³ Several scholars have indeed called for a better understanding of other-regarding preferences on the part of politicians, and the difference with the other-regarding preferences from the remainder of the population (e.g., Wilson, 2011, Hafner-Burton et al., 2013, Sheffer et al., 2018). Yet, only very few studies have endeavoured to tackle such an issue, and all of them are experimental political science studies relying on standard economic games. For instance, LeVeck et al. (2014) find evidence of pro-social behavior in an ultimatum game with policy and business elites: their sample shows higher rejection rates as second-movers in an ultimatum game as compared to both undergraduates and the general population. This can be interpreted as politicians' reservation level of "fairness" to be higher than the one of non-politicians. Similarly, Butler and Kousser (2015) find American legislators to cooperate more than a convenience student sample in a public good game, though they were also less responsive to social manipulations (deliberation and reputation). Finally, Enemark et al. (2016) exploit a discontinuity in the electoral success to show that office holders reciprocate more

¹³We acknowledge that there exists a semantical debate within the public administration literature concerning the interpretation of public service motivations as pro-social motivations (and conversely). Indeed some definitions of public service motivations explicitly link the two concepts: "Public service motivation can be defined as a general altruistic motivation to serve the interests of a community of people, a state, a nation, or humankind." (Rainey and Steinbauer, 1999). On the contrary, other definitions do not mention pro-sociality, which is the case of Perry and Wise (1990). Such a debate is obviously beyond the scope of this chapter, but we shall shortly precise why we believe that politicians may hold altruistic concerns.

in a trust game than non-office holders. All these findings suggest that politicians do actually hold pro-social concerns.¹⁴

However, pro-sociality does not necessarily reduce to altruism. Indeed, the literature on social or other-regarding preferences identifies many different ways of incorporating others into a decision-maker's utility (e.g., Camerer, 2011, Kerschbamer, 2015, Cooper and Kagel, 2016). In many cases, various motives overlap and disentangling them accurately requires careful experimental design or elaborate statistical analyses.¹⁵ In the remainder of this sub-section, we will specifically address the reasons why politicians may behave altruistically, and may differ from other groups in this regard.

First, the very large majority of politicians are volunteers, especially at the local level. As a consequence, they spend an important amount of time fulfilling public duties without receiving any salary. In other words, they are willing to sacrifice some personal benefits for the sake of others, and more generally for the sake of society. In their survey addressed to French mayors, Foucault and Bono (2020) find that the majority of respondents estimated spending between 10 to 40 hours per week on political activities. At the same time, many had a full-time (30,3%) or a part-time (15,3%) job, whereas the large majority was retired (45,7%). In addition, most respondents declared they perceive their public office as a "function" (69.4%) rather than a "job" (27.9%). To the extent that altruism has been found to be an important driver of the choice to volunteer (Burns et al., 2006, Carpenter and Myers, 2010), we shall expect politicians to hold altruistic views.

Second, the decision to participate into politics has been found to positively correlate with both field (Knack, 1992, Jankowski, 2007, Bali et al., 2020) and ex-

¹⁴More generally, there is a growing number of studies in the recent years that rely on experiments to investigate the question of measurement and representativeness of politicians' behavior. These include preferences under uncertainty (Fatas et al., 2007, Linde and Vis, 2017, Heß et al., 2018), time preferences (Sheffer et al., 2018) and honesty (Chaudhuri et al., 2020).

¹⁵For example, in a trust game a subject (the "trustor") is endowed with an amount of money to be shared with another subject (the "trustee") during a first phase. Any amount sent by the trustor is tripled. During a second phase, the trustee may return (part of) what he received to the trustor. The trustor's behavior is generally conceived as "trust" whereas the trustee's behavior is conceived as "reciprocity". Yet, both behaviors actually incorporate other motives than pure trust and reciprocity, namely unconditional motives (such as altruism or inequality-aversion). One way to control for such unconditional motives is to differentiate behavior in the trust game with behavior in a similar game when conditional motives have been neutralized (Cox, 2004).

perimental (Fowler, 2006, Fowler and Kam, 2007, Robalo et al., 2017, Kam et al., 2019) measures of altruism. Though the definition of political participation in these studies mostly boils down to the decision to vote¹⁶, it is arguable that the correlation with altruistic preferences extends to other types of political engagements (Jankowski, 2019). Indeed, the rationale for why other-regarding preferences increase the likelihood to vote is that voting ultimately affects the policies that will be implemented and which will have an impact on the whole population (Jankowski, 2007). Yet, one more direct way of affecting implemented policies would precisely be to conceive and enact them. In such a case, altruistic citizens would self-select into political candidacy. Furthermore, voters may appraise candidates holding other-regarding preferences as an insurance mechanism against policies moving away from their preferences. Indeed, one way for democracies to cope with the risk that the will of the people may be ignored would be to select political leaders that prioritize this will¹⁷, which entails a concern towards others (Gulzar and Khan, 2021). Such a type of character has however been mostly neglected by the literature on political selection.¹⁸ To the best of our knowledge, Bernheim and Kartik (2014) present the only theoretical example that considers altruistic motivations within a model of political selection. They lay the emphasis on "public-spiritedness", which relates to a concern towards others (altruism), and find an U-shaped relationship: only citizens with extreme values of public spirit sign up for political candidacy.¹⁹

Third, there also exists experimental evidence that politicians tend to behave pro-socially (LeVeck et al., 2014, Butler and Kousser, 2015, Enemark et al., 2016).

¹⁶More generally, the assumption that citizens hold other-regarding preferences (and foremost altruistic preferences) has been advanced by many scholars as a theoretically-valid and empirically-sound explanation to the voting paradox (among others: Edlin et al., 2007, Jankowski, 2007, 2019, Evren, 2012). The paradox stems from the fact that many people vote despite the cost of doing so outweighs by far the benefits due to the extremely-small probability of influencing the election results. Assuming people care about others' well-being resolves the paradox without dismissing the rational choice framework.

¹⁷As noted by Braendle (2016): "Conceptually, one might think of a "good" politician as being one who implements the policy that would be preferred by the median voter or the policy that would be chosen behind a veil of uncertainty".

¹⁸So far, such a literature has mainly focused on politicians' ability/competence or corruption/honesty while alternative qualities have mostly been ignored. See Braendle (2016), Dal Bó and Finan (2018) or Gulzar (2021) for recent surveys.

¹⁹Beniers and Dur (2007) also propose a framework where politicians differ in "Unselfishness" and emphasize the importance of "political culture" (i.e., other politicians' motivations), which may generate a self-sustaining opportunistic equilibrium. However, they do not endogenise political selection.

In particular, Vollan et al. (2020) make local Namibian leaders participate in binary allocation games and find that they behave more generously than villagers. Similarly, Chaudhuri et al. (2020) find that local Indian politicians send more in a standard dictator game than ordinary citizens, though the effect fades away with experience in political decision-making.²⁰

2.2 Equality/Efficiency tradeoff

The Equality/Efficiency dilemma corresponds to a long-lasting debate in political philosophy concerning the justice criterion that should matter in policy evaluation.²¹ On the one hand, utilitarian theories of justice posit that maximizing efficiency, defined as the sum of individual utilities, is the only legitimate moral target for societies, no matter the resulting inequalities. On the other hand, deontological theories of justice (e.g., Rawls, 1971) underline that neglecting how resources are ultimately distributed likely violates some morally superior criterion, which warrants infringing on an efficient allocation of resources. These two competing views had a long-lasting influence on the welfarist approach to governments' decisions.

Aside from these normative views, there also exists a long tradition of positive studies which incorporate a dilemma between minimizing income inequalities (through redistribution) and maximizing efficiency (through increasing total income). Early theoretical contributions that assume distributional concerns on the part of governments within a political economy framework include Behrman and Craig (1987) on public expenditures or Dixit and Londregan (1998) on redistributive policies.²² These initial models have been used to derive empirical tests based

²⁰A similar non-linear relationship between morally-reprehensible behavior and experience in policy-making is underlined by the authors for honesty. Also see Janezic and Gallego (2020).

²¹The traditional way of referring to this dilemma implies pitting equity against efficiency. However, "equity" is a loaded term in the distributional justice literature which may connect to many different motives. For instance, Konow (2003) identifies "equity" with a concern about individuals' responsibility for their actions (e.g., whether they are derived from choices or chances). In his denomination, such a motive is different from the "equality" motive which relates to reducing inequalities. We consequently chose a more direct terminology which slightly deviates from the standard practice.

²²Such a class of models builds upon a larger set of theoretical contributions which emphasizes the importance of policy motivations (or "ideology") alongside standard vote-maximizing motivations in explaining policy patterns (see most notably Wittman, 1977, Calvert, 1985). However, the ideology space encompasses a wide array of topics and does not necessarily overlap the set of distributional preferences.

on reduced-form econometric specifications to investigate both the weights placed on the Equality/Efficiency tradeoff and the relative importance of such tradeoff with respect to political factors akin to strategic self-interested decisions (e.g., maximizing votes). The tests have used data on safety expenditures (Behrman and Craig, 1987, Craig, 1987), education expenditures (Behrman and Birdsall, 1988, Gershberg and Schuermann, 2001), or infrastructure investments (Castells and Solé-Ollé, 2005, Bel and Fageda, 2009, Kemmerling and Stephan, 2010, Solé-Ollé, 2013). Although the evidence regarding both the influence of political factors as well as the importance of Equality versus Efficiency is rather mixed²³, these studies lay the emphasis on the empirical relevance of assuming that distributional concerns matter to policy-making. Contrary to these studies, we evaluate the importance for politicians of the Equality/Efficiency tradeoff relative to other motivations (Selfishness v. Altruism) at the individual level by directly observing decisions in a controlled environment. Consequently, we do not need to make assumptions on how politicians' individual distributional preferences ultimately affect decisions (i.e., policies), our analysis does not rely on reduced-form specifications and we take into account the possible heterogeneity of preferences instead of considering only average effects.

In addition, many findings from the literature in behavioral and experimental economics suggest that distributional concerns are likely to play an important role in political decisions.

First, experiments that basically replicate part of politicians' decision-making environment emphasize the importance of distributional preferences. For instance, Traub et al. (2009) find that subjects placed in a social planner position with known income distribution are more likely to be sensitive to a "comparative fairness" motive and select an equitable income distribution than social planners under the veil of ignorance. In a taxation/redistribution game that basically implements Okun (1975)'s metaphor, Durante et al. (2014) show the importance of both equality and efficiency considerations in explaining subjects' behavior, along with the traditional

²³For instance, a greater weight attributed by to Efficiency over Equality is found in Castells and Solé-Ollé (2005), Kemmerling and Stephan (2010) and Solé-Ollé (2013) whereas the converse is observed in Behrman and Craig (1987), Craig (1987) and Gershberg and Schuermann (2001), while Behrman and Birdsall (1988) emphasize equal weights and Bel and Fageda (2009) do not observe governments to care about such tradeoff. At the same time, political factors matter in all studies, though their importance varies.

self-interest and fairness views as to the origin of the distributed resources. Alekseev et al. (2021) underline that behavioral motives (such as efficiency-orientation) may help explain the flypaper effect, which is a classical conundrum in local public finance.²⁴ Contrary to these experiments, which rely on convenience student samples, we evaluate the specificity of actual politicians' distributional concerns as opposed to control groups, and in particular a student control group. Our study is thus a first step towards bridging the gap between standard experimental results and actual decisions made by politicians.

Second, a recent literature relying on both representative surveys and large-scale experiments underlines that distributional preferences matter to citizens' demand for various real-world policies. Using a representative survey from the US population as well as supplementary experiments, Stantcheva (2021) emphasizes that distributional concerns explain the support for tax schemes. Gniza et al. (2022) rely on a vignette study with German respondents to emphasize that both reducing inequalities and increasing total surplus affect the support for regional funding, depending on whether living or economic conditions are scrutinized. In an incentivized survey experiment in Germany, Kerschbamer and Müller (2020) find self-interested subjects to be less supportive of redistribution and immigration policies, whereas the converse is observed for altruistic, efficiency-oriented and equality-oriented subjects. Similar results are reported in Switzerland by Epper et al. (2020), who also emphasize that the reasons behind such support vary between efficiency-oriented and equality-oriented subjects.²⁵ Contrary to this literature, we focus on supply-

²⁴The flypaper effect is observed when the level of public expenditure is more responsive to an increase in lump-sum grants than to an increase in citizens' income, in contradiction with standard economic intuition according to which the effect should be the same (Hines and Thaler, 1995).

²⁵A different strand of the literature also examines the link between redistributive policies, or more generally the prevailing conception of the welfare state, and fairness ideals within the population (Alesina et al., 2004, Alesina and Angeletos, 2005, Alesina and Giuliano, 2011). This literature emphasizes that implemented policies crucially depend on what the population judges as fair. For instance, societies which attribute income inequalities to differences in effort (e.g., the United States) rather than differences in luck (e.g., France) are less likely to favor massive taxes-and-transfers schemes, and conversely. However, "fairness" in this literature is defined based on the sensitivity to the source of income inequalities (e.g., merit, luck or choice). Such a definition is somewhat connected to our conception of "distributional preferences" (e.g., the emphasis on egalitarianism, which implies treating inequalities as unjust and reducing all of them) and addressing how actual politicians trade-off these "fairness concerns" would certainly be fruitful from a political economy perspective. But it remains nevertheless conceptually different from the kind of motivational tradeoffs we are interested in. For an experimental investigation of such fairness ideals, see in particular Cappelen et al. (2007, 2010, 2013).

side agents (politicians), who are best placed to implement their own preferences, instead of demand-side agents (citizens).

There already exists some pointers as to how politicians would resolve the Equality/Efficiency dilemma in the context our experiment. For instance, Dawes et al. (2011) examine the link between distributional preferences and political participation using a similar design as ours with students. They observe that efficiency-oriented subjects are more likely to be politically active, contrary to other distributional types. They mention a possible rationale for such link based on Harsanyi (1975): equality-oriented individuals can only rarely achieve their goal through politics and may consequently prefer other activities (like charitable donation) because politics only rarely implements the kind of redistributive policies that they praise. Similarly, Fisman et al. (2015b) find elites to be more efficiency-oriented than the remainder of the population. Finally, Vollan et al. (2020) find that local leaders are less egalitarian than villagers.

2.3 Conjectures

To the best of our knowledge, we reviewed above all existing scholar work that included at some point an empirical analysis of distributional concerns on the part of politicians at the individual level. These sparse findings may serve as a basis to formulate some behavioral conjectures regarding how politicians may or not differ from both our control groups in the way they trade-off selfishness against altruism and equality against efficiency.

Regarding the Selfishness/Altruism tradeoff, we expect politicians to deviate from strict self-interest in at least some way. Indeed, observing a non-negligible non-selfish behavior in dictator games is a standard result in the experimental literature (Andreoni and Miller, 2002, Engel, 2011, Camerer, 2011, among many others). Besides, we expect to observe an empirical tension between selfishness and altruism within our target population, as both motivations appear *a priori* empirically valuable in explaining politicians' behavior. Assessing which motivation dominates is precisely one objective of our study.

Hypothesis 1 (Politicians and the Selfishness/Altruism tradeoff, H^1):
Politicians deviate at least few times from strictly selfish behavior.

At the same time, the existing evidence comparing politicians with non-politicians on this issue is rather mixed. We therefore cannot formulate a clear working hypothesis and consider that there is *a priori* no reason to consider the difference between politicians and a comparable control group to lean more towards one direction.

Hypothesis 2 (group comparison in Selfishness/Altruism tradeoff, H^2): *Politicians do not differ significantly from a comparable control group in terms of Selfishness/Altruism.*

However, there also exists multiple evidence according to which students or young people behave more selfishly than other fringes of the population in distribution games (e.g., Bellemare et al., 2008, Engel, 2011, Fisman et al., 2017, Frigau et al., 2019). Coupling this stylized fact with our second hypothesis, we can derive an hypothesis with respect to the comparison between Politicians and Students.

Hypothesis 3 (group comparison in Selfishness/Altruism tradeoff, H^3): *Politicians are more concerned about others than Students.*

Regarding the Equality/Efficiency tradeoff, existing sparse findings point towards assuming that Politicians are likely to value more efficiency over equality and are more likely to do so than Non-Students.

Hypothesis 4 (Politicians and the Equality/Efficiency tradeoff, H^4): *Politicians favor efficiency over equality overall.*

Hypothesis 5 (group comparison in Equality/Efficiency tradeoff, H^5): *Politicians display a greater concern for efficiency over equality than a comparable control group.*

However, we cannot formulate a clear hypothesis with respect to the difference between Politicians and Students. Indeed, students or young people have also been shown to favor efficiency over equality (e.g., Bellemare et al., 2008, 2011, Fisman et al., 2017, Müller, 2019, Kerschbamer and Müller, 2020). The distributional preferences of both Politicians and Students are thus expected to be oriented towards efficiency, and there does not seem to be any valid reason to assume *a priori* that one group should value efficiency more than the other.

Hypothesis 6 (group comparison in Equality/Efficiency tradeoff, H^6):
Politicians do not necessarily differ from Students in the way they trade-off equality and efficiency.

Finally, we do not make a strong statement as regards a possible difference between Politicians from the Lab and Politicians from the Field. From a theoretical perspective, we could expect the costs to come into the laboratory to be higher than the costs to participate in field sessions. As a consequence, we could expect a higher proportion of selfish types among Politicians from the Lab, because only they would be sufficiently interested to participate despite the costs. At the same time, reducing the willingness to participate in experiments to this sole extrinsic motivation would probably be too narrow. It is indeed plausible that participants wanted to help research or were interested in the basic fact that science showed some interest in them. In such a case the relationship with self-interest is not clear anymore, and may even reverse. Furthermore, the cost dimension is not the only dimension to consider. Indeed, Politicians from the Field had a higher probability than lab participants to have already been acquainted with other participants in the same session because they took part in the same intermunicipal government. Although anonymity was strictly guaranteed throughout the experiment, there is a possibility that such a feature interacted with observed behavior anyway. However, the nature of such interaction would rely upon existing relationships between participants and may not necessarily imply a spike in altruistic behavior (e.g., if participants dislike each other). In addition, other features of the experimental procedure were different between the Lab and the Field (see the sub-section 3.1 from the chapter 1), so that the overall effect on observed behavior we may expect is not straightforward.

Our study differs in several respects from the above-mentioned papers involving actual politicians participating in laboratory experiments (and the corresponding ones from the previous sub-section). First, our methodology allows to finely disentangle two qualitatively different tradeoffs: the Selfishness/Altruism tradeoff and the Equality/Efficiency tradeoff. These tradeoffs are typically intertwined in simple games with few repetitions. Besides, as we expose each subject to many decisions, we are able to disentangle both tradeoffs at the individual level and thus to account for individual heterogeneity in distributional concerns instead of considering aggregated behavior only. Second, the high number of individual decisions allows

to implement revealed preference tests in order to screen out inconsistent behavior that may bias the results. These two features are absent in all the above-mentioned studies, except for Fisman et al. (2015b). However, the latter focuses on "elites", as defined by education and/or income, whereas we measure the distributional preferences of actual local politicians. Third, we focus on local politicians in a developed country (France), whereas all existing studies involving actual politicians occurred in developing countries. Arguably, the political institutions in developed countries are likely to differ from their counterparts in developing countries, which may translate into different preferences of the political elite. Fourth, we compare our sample of politicians with a counterfactual control group based on available observables. This aims at reducing the confounding effects of some variables known to impact distributional concerns (such as age or gender). On the contrary, existing studies compare politicians with citizens taken from the general population with no specific sample restrictions.

Our study also documents the growing literature on comparing convenience experimental samples (students) and other populations, which addresses the issue of the generalizability of the results obtained by experimental studies (see among others Bellemare and Kröger, 2007, Carpenter et al., 2008, Choi et al., 2014, Cappelen et al., 2015, Frigau et al., 2019). Relatedly, our study allows to question the replicability of existing results in a different context. Indeed, most studies relying on a similar methodology as ours draw their sample from the American population (including Andreoni and Miller, 2002, Fisman et al., 2007, 2015a,b, 2017, Dawes et al., 2011, Jakiela, 2013). Yet, distributional concerns have been shown to vary across societies (see among others Alesina and Giuliano, 2011, Cappelen et al., 2013, Jakiela, 2015, Falk et al., 2018, Almås et al., 2020).²⁶ Finally, our experiment connects with the literature comparing results between laboratory and field settings (Eckel and Grossman, 2000, Güth et al., 2007, Benz and Meier, 2008).

²⁶This is all the more relevant as some subsequent studies also rely on datasets supplied by the early studies. For instance, Heufer (2013) and Heufer and Hjertstrand (2019) test for homothetic preferences based on the data from the experiments of Fisman et al. (2007) and Choi et al. (2007).

3 Experimental design

3.1 Subject pools and procedures

Our experiment recruits from three distinct groups within the French population: local French politicians (henceforth "Pol"), people from the general population who resemble politicians but do not hold a local term (henceforth Non-student Control, "NSC") and university students (henceforth Student Control, "SC"). Our sample of politicians (139 subjects) is composed of two sub-groups: one group came into the lab at the LABEX-EM of the University of Rennes I (Politicians from the Lab, 67 subjects) and the other group participated in field sessions on tablet computers (Politicians from the Field, 72 subjects). Participants to the laboratory sessions were recruited thanks to the support of the Association des Maires de France (AMF), which is the biggest French association of mayors.²⁷ Field sessions were organized directly within some intermunicipal governments in Brittany. Both our control groups participated in laboratory sessions at the LABEX-EM. Non-Students were recruited through several channels: announcements in a regional newspaper, advertizements broadcasted through the university's networks, posters in the city of Rennes, etc. (91 subjects). Interested people had to fill in an online form which enabled us to screen candidates based on some socio-demographic characteristics (gender, age and occupation) in order for that group to ultimately look like our sample of politicians in term of these characteristics. Students were recruited through ORSEE (Greiner, 2015) within a pool of students from Rennes (123 subjects).

The same procedures applied to every session: subjects were greeted and explained the most basic rules of an experiment ; a ball-in-the-urn process randomly attributed them stations in the laboratory ; they listened to the pre-recorded instructions all together and could also read it ; they played the modified dictator game programmed in z-Tree (Fischbacher, 2007) ; they answered a short socio-demographic questionnaire ; and finally they individually received their paycheck.

²⁷Participation was however not restricted to mayors. Indeed, within the French local system of representation, decisions are subject to a vote involving the other local councillors. The latter's preferences thus matter in the decision-making process. Furthermore, it is an open question to assess whether distributional preferences rely upon the degree of involvement within the local government.

The game lasted 45 minutes on average and all amounts were expressed in Experimental Currency Units, with $100 \text{ ECU} = 7\text{€}$.

The game consisted in 30 decisions per subject. For each decision, the subject was both a dictator and a recipient: he decided as a dictator for the subject with whom he had been randomly matched and he was also the recipient in this other subject's decision. The dictator-recipient pair changed for each decision. No feedback was implemented across decisions and the strict anonymity of decisions between participants was underlined in the instructions. At the end of the experiment, each subject's payoff for the entire game was computed as the sum of i) the ECU she kept from one randomly-selected decision among the decisions she made as a dictator, and ii) the ECU she received from one randomly-selected decision among the decisions other subjects made for her when she is a recipient. Each subject became aware of her payoff only after the experiment was completed. Such payoff was composed of a 6€ flat fee and an incentive part depending on the decisions made during the experiment. Specifically, we selected only a third of subjects to receive the incentive part at the end of each session. The remaining two thirds were only given the flat fee.²⁸

The data collected for the present chapter has also been used in the previous chapter, which focuses on testing the consistency of politicians' decisions with a rational choice framework. Our major finding is that most data can be treated as utility-generated, in line with the remainder of the literature which focuses on non-politician samples. In the current chapter, we make use of such finding to proceed with assuming and estimating a CES utility function (see section 4.2) in order to recover distributional preferences at the individual level, in the spirit of Fisman et al. (2007, 2015a,b, 2017). In addition, we refer the interested reader to the sub-section

²⁸Selecting only a subset of subjects for payment is known as the Between-subject Random Incentive System (BRIS). When combined with paying only one decision out of several (the Within-subject Random Incentive System, WRIS), such a payment scheme is known as the Hybrid Random Incentive System (HRIS: Baltussen et al., 2012). Such a specific payment scheme was designed to mitigate potential recruitment or behavioral biases relative to our target group (politicians). Obviously, it may modify observed behavior in the game due to incentives' dilution. Yet, existing empirical evidence on the impact of such aspect of the payment scheme concludes that it remains small, at least as far as simple distribution games are concerned (Charness et al., 2016, Clot et al., 2018). We will return to this specific issue in the discussion part below.

3.1 from the chapter 1 for a detailed description of the experimental procedures as well as a compositional analysis of our subject pools.

3.2 The modified dictator game

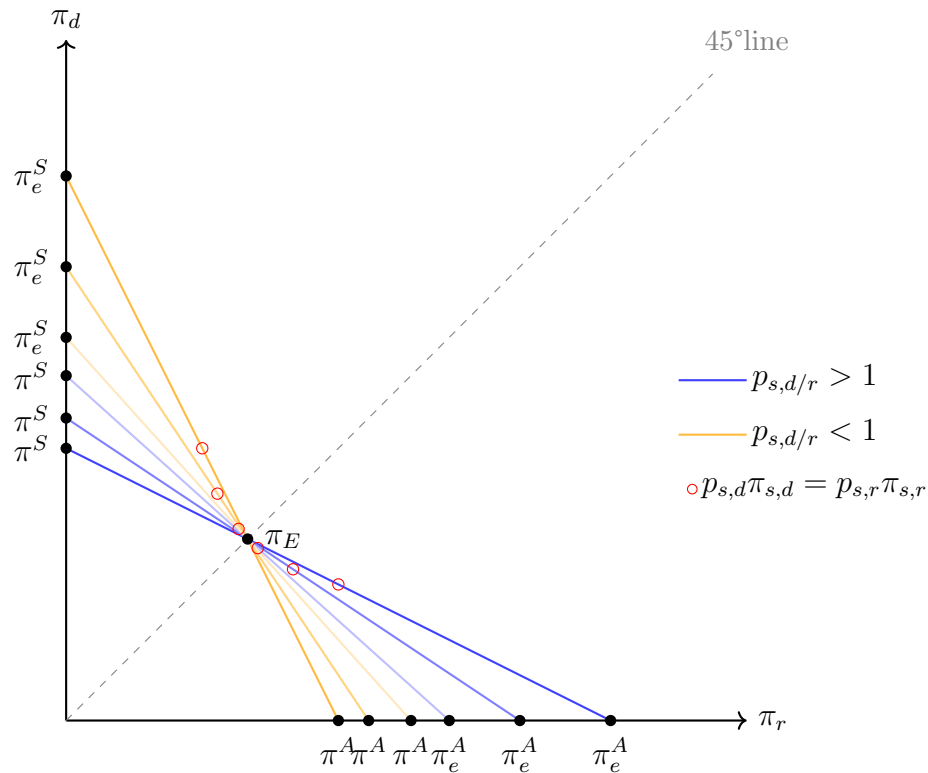
Our experimental game consists in the modified version of the dictator game with varying relative prices of giving/keeping the endowment, as proposed by Andreoni and Miller (2002).²⁹ Formally, each modified dictator game i asks each subject s as dictator d to select an ECU allocation $(\pi_{s,d}^i, \pi_{s,r}^i) \geq 0$ between herself and a recipient r at corresponding prices $(p_{s,d}^i, p_{s,r}^i)$ under the (normalized) budget constraint $B_s^i = p_{s,d}^i \pi_{s,d}^i + p_{s,r}^i \pi_{s,r}^i = 1$.³⁰ Following Fisman et al. (2007), the game is introduced to subjects graphically as a choice from a two-dimensional budget line. Each dot on the line represents an allocation $(\pi_{s,d}, \pi_{s,r})$ of ECU for the dictator (y -axis) and for another randomly-selected subject in the room acting as recipient (x -axis). Subjects are then instructed to click on the desired dot, without any time limitation. The payoffs corresponding to each allocation are directly readable in a help frame on the right of the line as well as in frames below the x -axis and on left of the y -axis. We expose each subject to thirty such decisions. All of the thirty budget lines are randomly selected from the set of budget lines which meet two criteria: *i*) a budget line cannot intersect any axis at more than 100 ECU, and *ii*) a budget line cannot intersect both axes below 50 ECU. Each decision problem is thus independent from the others (within and between subjects). The relative price can be computed ex-post as $p_{s,d/r}^i = p_{s,d}^i / p_{s,r}^i = \bar{\pi}_{s,r}^i / \bar{\pi}_{s,d}^i$, with $\bar{\pi}_{s,d}^i$ and $\bar{\pi}_{s,r}^i$ the i -th budget line endpoints. In other words $p_{s,d/r}^i$ corresponds to the inverse of the slope of the budget line: the steeper it is, the lower $p_{s,d/r}^i$ is.

The figure 2.1 illustrates possible budget lines and emphasizes several theoretical

²⁹In the traditional version of the dictator game (Forsythe et al., 1994), the relative price is constrained to one. This makes it impossible to disentangle the different behavioral motives behind allocation decisions (in particular a concern for others and some taste for equality or efficiency).

³⁰We will henceforth denote $p_{s,d}^i \pi_{s,d}^i / B_s^i$ the hold rate and $p_{s,r}^i \pi_{s,r}^i / B_s^i$ the pass rate of subject s in decision i .

Figure 2.1: Examples of budget lines



Note: The black dots indicate the theoretical prediction for a subject who strictly abides by one of the four motives we are interested in, namely selfishness, altruism, equality and efficiency. Each red circle reflects the centroid of the corresponding budget line, that is the allocation for which the pass rate equals the hold rate. All together, the red circles represent a typical pattern of interior choices (the behavior of a subject with a Cobb-Douglas utility function) which can serve as a reference point to assess other types of intermediate behavior.

insights for the underlying behavioral motives we scrutinize.³¹ The black dots represent theoretical predictions with only one "pure" distributional preference in mind. The red circles represent the centroids of the budget lines, that is the allocation which equalizes the hold and the pass rates. The position of the chosen allocation with respect to the centroid is therefore informative of which role the decision-maker

³¹For ease of exposure, all budget lines cross the same dot π_E , which maximizes the utility of an equality-oriented dictator. In practice, the budget lines of a given subject did not necessarily cross at the strict equality prediction (and could even not cross at all).

values most in a given situation: a choice on the left (resp. right) of the centroid indicates that more weight is placed on the dictator (resp. recipient).³²

A strictly selfish dictator would maximize $u(\pi_{s,d}^i, \pi_{s,r}^i) = \pi_{s,d}^i$ and systematically choose the allocations lying at the intersection of the budget line and the y -axis line, that is allocations π_k^S ($k \in (\emptyset, e)$).

A strictly altruistic dictator would maximize $u(\pi_{s,d}^i, \pi_{s,r}^i) = \pi_{s,r}^i$ and systematically choose the allocations lying at the intersection of the budget line and the x -axis line, that is allocations π_k^A ($k \in (\emptyset, e)$).

As a consequence, strictly selfish and strictly altruistic behaviors lead to opposite predictions, so that subjects are facing a tradeoff between selfishness and altruism. Intuitively, the average position of the chosen allocation with respect to the centroid over all the thirty budget lines thus indicates how the subject trades off selfishness and altruism.

A strictly equality-oriented dictator would strive to reduce payoff differences between the two roles and thus grant both with the same amount. Such a subject would thus maximize $u(\pi_{s,d}^i, \pi_{s,r}^i) = \min(\pi_{s,d}^i, \pi_{s,r}^i)$ and systematically choose the allocations lying at the intersection of the budget line and the evenly distribution line, that is allocation π_E .

A strictly efficiency-oriented dictator would strive to maximize the sum of payoffs, which in our context means to give the available budget to the role whose tokens are the cheapest. Consequently, such a subject would maximize $u(\pi_{s,d}^i, \pi_{s,r}^i) = \pi_{s,d}^i + \pi_{s,r}^i$ and systematically choose the allocations lying at the intersection of the budget line and one of the axes. Such a behavior will thus depend upon the relative price: when $p_{s,d/r} > 1$ the recipient is the role that can earn the highest payoff, whereas when $p_{s,d/r} < 1$ the opposite turns true. Ultimately such a behavior corresponds to allocations π_e^k ($k \in (S, A)$).

More generally, the reaction to relative price changes is indicative of how the subject trades off equality and efficiency. An equality-oriented dictator will be willing to equalize both subjects' payoffs, hence spending a larger endowment share on

³²The pattern displayed by the red circles is a typical pattern of interior choices corresponding to a Cobb-Douglas utility function with equal weights placed on both roles: $u(\pi_{s,d}^i, \pi_{s,r}^i) = (\pi_{s,d}^i \pi_{s,r}^i)^{1/2}$.

the role whose tokens are the most expensive (i.e., facing the highest price). Favoring equality thus requires less and less weight to be placed on the recipient as $p_{s,d/r}$ increases: the gap between the strict equality prediction and the centroid shrinks when $p_{s,d/r} < 1$ and grows when $p_{s,d/r} > 1$ as the relative price increases. On the contrary, an efficiency-oriented dictator will be willing to maximize total income, hence spending a larger endowment share on the role whose tokens are the cheapest. Favoring efficiency thus requires less and less weight placed on the dictator as $p_{s,d/r}$ increases.³³ Ultimately, minimizing income inequalities and maximizing total income predict opposite behavior with respect to price changes: equality (efficiency) concerns imply spending less (more) on the recipient as $p_{s,d/r}$ rises.

Resorting to such modified dictator games has several advantages, as opposed to other possible experimental designs. First, by exposing each subject to multiple linear budget constraints with varying relative prices, our experiment allows to resort to standard revealed preference tools in order to test for individual data's compatibility with utility maximization.³⁴ Second, a continuous choice over a budget line involves more statistical information than a discrete choice within a pre-defined list of allocations. It thus provides a more accurate description of subjects' preferences. In particular, it enables the use of econometric modelling to carefully disentangle between different components of observed behavior: how subjects trade off selfishness versus altruism on the one hand and equality versus efficiency on the other hand. Third, it provides the researcher with numerous individual observations, which makes it possible to individually specify the structural model's error term and to estimate subject-level parameters. On the contrary, most existing studies focus on aggregate measures of behavior based on pooling individual data and con-

³³Graphically displaying this corollary for efficiency would have required a figure with all budget lines crossing at the same strict prediction for efficiency (hence with one prediction for strict equality per budget line). Such a figure would indeed have shown that the gap between the strict efficiency prediction and the centroid grows when $p_{s,d/r} < 1$ and shrinks when $p_{s,d/r} > 1$ as the relative price increases.

³⁴We refer the interested reader more specifically to the list of benefits related to the revealed preference analysis that is reported in the section 3.2 from the chapter 1 for more information.

sequently fail to finely account for individual heterogeneity.³⁵ Fourth, the graphical interface is arguably intuitive and does not require any computation on the subjects' part, which facilitates their understanding of the game. Considering that our target population (politicians) is very unlikely to have prior experience in economic experiments, such a feature is highly appealing.

4 Distributional preferences analysis

4.1 Revealed preference analysis: consistency with homothetic preferences

In order to carefully disentangle the underlying distributional preferences behind observed behavior, the methodology we adopt consists in estimating a structural homothetic utility function.³⁶ One major requirement for such methodology to apply is that behavior is indeed compatible with a homothetic utility maximization framework. Most studies relying on the same methodology only consider testing for the compatibility with utility maximization by implementing some relaxed version of the Generalized Axiom of Revealed Preference (GARP) test (Afriat, 1967, Varian, 1982). Using this test to narrow the analysis to GARP-consistent subjects, they proceed with modelling behavior based on a CES utility function. However, the GARP test does not provide any information as to the functional form of the underlying utility function (although a CES function is homothetic). Building on recent advances in the revealed preference literature (in particular Heufer and Hjertstrand, 2019), we go one step further and test for homothetic utility maximization. To the best of our knowledge we are the first study to implement such techniques on novel experimental data: even Heufer and Hjertstrand (2019) compute their measures using existing datasets.

³⁵We acknowledge that the bi-dimensionality of our experimental game prevents from analyzing some other types of distributional preferences. In particular, reducing payoff differences and targetting the least well-off individual (maximin preferences) are conceptually different and have been shown to differently motivate experimental subjects (see for instance Engelmann and Strobel, 2004). Yet, both distributional preferences are merged into the equality prediction in our setting.

³⁶A function is said to be "homothetic" if it is a monotonic transformation of another function that is homogeneous of degree 1. In other words, a homothetic utility maximizer that is willing to buy π_s^i at prices p_s^i will also prefer $\lambda\pi_s^i$ when prices are constant but income is multiplied by λ .

Classical revealed preference theory provides a straightforward test based on the seminal theorem by Varian (1983): given a finite collection of budget sets, observed decisions can be rationalized by a well-behaved (piecewise linear, continuous, monotonically increasing and concave) homothetic utility function if and only if they comply with the Homothetic Axiom of Revealed Preference (HARP). HARP is satisfied if, for all distinct choices of indices i, j, \dots, l we observe $(p_s^i \pi_s^j)(p_s^j \pi_s^k) \dots (p_s^l \pi_s^i) \geq B_s^i B_s^j \dots B_s^l = 1$.

Given that testing HARP only indicates whether the subject is compatible with homothetic preferences or not, which does not account for likely mistakes of small intensity, we rather focus on several continuous measures of closeness to homothetic utility maximization. The main one is the Homothetic Efficiency Index (HEI, see Heufer, 2013, Heufer and Hjertstrand, 2019), which quantifies the amount by which all budget lines must be shifted in order to remove all HARP violations. The difference between the HEI and one can be interpreted as the fraction of the endowment that the subject is wasting by making choices that are inconsistent with homothetic utility maximization. Consequently, the closer the HEI is to one, the closer the subject is to behave as if he is maximizing a homothetic utility function. However, the HEI neglects the importance of all violations except the biggest one. In particular, it considers repeated violations by small amounts as closer to homothetic preferences than an unique big violation, which is debatable, and generally it does not provide any information as to the number of violations or their associated inefficiencies. We therefore compute two additional indices: the Homothetic Houtman-Maks Index (H²MI, see Heufer and Hjertstrand, 2015) and the Homothetic Efficiency Vector (HEV, see Heufer and Hjertstrand, 2019). The H²MI indicates the biggest set of observations that is strictly consistent with homothetic utility maximization. The HEV is a vector that emphasizes how much perturbation must be minimally made on each budget in order to remove all HARP violations overall.

For the sake of parcimony, we present here only a summary of our results on revealed preference tests (a detailed presentation is exposed in appendix B.1). We find that assuming homothetic preferences impose stringent restrictions on observed behavior. Specifically, no subject is strictly consistent with HARP: Politicians and Non-Students waste approximately 40% of their endowment by not abiding by HARP (HEI of 0.6), whereas Students are closer to homothetic utility maximization (HEI of 0.75). Moreover, making individual datasets comply with HARP

requires removing on average about 20 observations in all groups (Pol, Student Control and Non-Student Control). Nevertheless, the results on the HEV show that most subjects ultimately fall close to homothetic utility maximization: despite few big violations, most efficiency levels in the vector are very close to one. In particular, the average mean of homothetic efficiency levels in the HEV is above 0.95 for all groups: the average behavior is 95% as efficient as maximizing a homothetic utility function. The same applies to the average first quartile of homothetic efficiency levels: for all groups, more than 3 out of 4 observations within a subject's dataset are associated with violations lower than 5% of the endowment on average. In addition, we compare each of our subjects' closeness to homothetic preferences to a counterfactual simulated dataset composed of hundred uniform-random decision-makers deciding out of the same 30 budget lines. For each simulated subject we compute the three homothetic indices and subtract them to the indices from the corresponding actual subject. This provides a way to gauge the predictive success of the homothetic utility maximization model, that is its ability to make sense of observed data as opposed to an alternative random choice model (Selten, 1991, Dean and Martin, 2016). Overall, the former outperforms the latter by far in our data. For instance, the HEI from our subjects is 35 percentage points higher than the HEI from simulated subjects, while the average element in the HEV is 8.4 percentage points higher. In other words, homothecity appears quite demanding for our subjects but they still appear much more consistent with it than if they had chosen at random along the budget lines. Ultimately, sticking to the CES utility function for the many subjects who display only marginal violations of homothecity still sounds reasonable.³⁷

4.2 Theoretical framework and empirical strategy

The above revealed preference analysis shows that most of our data can be rationalized by a well-behaved homothetic utility function. We consequently build upon such a result to impose more structure on the data by assuming that each subject's

³⁷A detailed presentation of the three homothetic indices' computation is exposed in appendix B.1. Such appendix also contains a more detailed empirical analysis on between-groups comparisons as well as on the predictive success of the model.

decisions result from maximizing a CES utility function.³⁸ Such an assumption will make it possible to structurally estimate individual parameters for the two trade-offs under scrutiny and to make group comparisons based on these more precise measures of underlying preferences.

In our setting where the subject s is asked under decision $i = 1, \dots, 30$ to divide an endowment between herself and another subject at relative price $p_{s,d/r}^i = p_{s,d}^i/p_{s,r}^i$, yielding payoffs $\pi_{s,d}^i$ (for the dictator) and $\pi_{s,r}^i$ (for the recipient), the CES function takes the following form:

$$u(\pi_{s,d}^i, \pi_{s,r}^i) = (\alpha_s(\pi_{s,d}^i)^{\rho_s} + (1 - \alpha_s)(\pi_{s,r}^i)^{\rho_s})^{1/\rho_s} \quad (2.1)$$

where $\alpha_s \in [0, 1]$ denotes the sensitivity to Selfishness/Altruism (higher α_s means more weight attached to favoring self) and $\rho_s \in]-\infty, 1]$ characterizes how the individual trades off equality and efficiency (and corresponds to the curvature of the indifference curves.³⁹)

The primary benefit from assuming a CES functional form lies on its ability to accurately disentangle the Selfishness/Altruism and the Equality/Efficiency trade-offs. Indeed, it spans a wide range of well-behaved utility functions and incorporates the canonical types of preferences we are interested in as special cases (Mas-Colell et al., 1995).

First, the boundaries of α_s define opposite preferences types in terms of the Selfishness/Altruism tradeoff: for any arbitrary ρ_s , when $\alpha_s = 1$ we obtain purely

³⁸According to Varian (1982)'s version of Afriat (1967)'s theorem, consistency with GARP implies that the utility function rationalizing the data can be chosen to be increasing, continuous and concave (and homogeneous of degree zero in the case of two goods). Further assuming homotheticity and separability, this function must belong to the Constant Elasticity of Substitution (CES) family. Such a functional form is rampant in demand analysis and is used for instance by Andreoni and Miller (2002), Fisman et al. (2007, 2015a,b, 2017), Jakiela (2013), and Müller (2019) (among others). Alternative formulations of utility in a similar purely distributional context generally rely on a (piece-wise) linearity assumption and typically build on pioneer models from either Fehr and Schmidt (1999) or Charness and Rabin (2002). The CES utility function is also regularly used in the political economy literature that incorporates the Equality/Efficiency dilemma in the maximization program of the policy-maker (e.g Behrman and Craig, 1987, Dixit and Londregan, 1998, Castells and Solé-Ollé, 2005).

³⁹On the relationship between indifference curves and the parametrization of the utility function, we refer the interested reader to the graphical representation proposed by Fisman et al. (2015b, 2017).

selfish preferences $u(\pi_{s,d}^i, \pi_{s,r}^i) = \pi_{s,d}^i$, whereas when $\alpha_s = 0$ we obtain purely altruistic preferences $u(\pi_{s,d}^i, \pi_{s,r}^i) = \pi_{s,r}^i$.

Second, given fair-minded preferences (i.e., $\alpha_s = 1/2$), the boundaries of ρ_s define opposite preferences types in terms of the Equality/Efficiency tradeoff: Rawlsian (or Leontief) preferences versus utilitarian preferences.⁴⁰ Indeed when $\rho_s \rightarrow -\infty$, the subject's utility converges on the maximin utility $u(\pi_{s,d}^i, \pi_{s,r}^i) = \min(\pi_{s,d}^i, \pi_{s,r}^i)$, meaning she will seek to perfectly equalize payoffs and choose the dot at the intersection of the budget line and of the evenly distribution line (45°). On the contrary when $\rho_s \rightarrow 1$, the subject's utility converges on the sum of payoffs $u(\pi_{s,d}^i, \pi_{s,r}^i) = \pi_{s,d}^i + \pi_{s,r}^i$, meaning she will seek to increase total payoffs and choose the dot at the intersection of the budget line and of one of the two axis (depending on the position of $p_{s,d/r}^i$ with respect to 1).

Third, when there is no bias towards either equality or efficiency, that is when $\rho_s \rightarrow 0$, the subject's utility approaches the Cobb-Douglas utility function $u(\pi_{s,d}^i, \pi_{s,r}^i) = (\pi_{s,d}^i)^{\alpha_s} (\pi_{s,r}^i)^{1-\alpha_s}$. Consequently, the optimal endowment shares $(p_{s,k}^i \pi_{s,k}^i)^*$ ($k \in (d, r)$) are constant across all prices and respectively equal α_s and $1 - \alpha_s$. In particular, when the subject is fair-minded and $\rho_s \rightarrow 0$, the utility function approaches the Nash product $u(\pi_{s,d}^i, \pi_{s,r}^i) = \pi_{s,d}^i \pi_{s,r}^i$ which strikes a balance between maximizing the sum of payoffs and reducing payoff inequalities.

Maximizing (2.1) under the budget constraint $p_{s,d}^i \pi_{s,d}^i + p_{s,r}^i \pi_{s,r}^i = 1$ yields the optimal endowment share spent on self (the dictator):

$$(p_{s,d}^i \pi_{s,d}^i)^* = \frac{\gamma_s}{(p_{s,d/r}^i)^{\kappa_s} + \gamma_s} \quad (2.2)$$

where $\gamma_s = (\alpha_s / (1 - \alpha_s))^{1/(1-\rho_s)}$ and $\kappa_s = \rho_s / (1 - \rho_s)$

The sign of ρ_s indicates whether the subject's distributional preferences are weighted towards either equality or efficiency. When $\rho_s > 0$, the optimal endowment share spent on self $(p_{s,d}^i \pi_{s,d}^i)^*$ decreases in reaction to an increase in the relative price of keeping the endowment $p_{s,d/r}^i$. Such a behavior points towards efficiency-oriented preferences: less endowment is spent on the dictator when keeping money

⁴⁰Such a dichotomy is also known as the perfect complements/perfect substitutes dichotomy used for comparing goods or production factors.

is relatively more costly (i.e., increasing her share of endowment would decrease the sum of payoffs). Conversely, any $\rho_s < 0$ is indicative of an equality-oriented subject since $(p_{s,d}^i \pi_{s,d}^i)^*$ increases when $p_{s,d/r}^i$ increases: more endowment is spent on the dictator when keeping money is relatively more costly (i.e., increasing the recipient's share of endowment would increase payoff inequalities). Such a relationship between distributional preferences and the sign of ρ_s provides a reduced-form test of the importance of a given type of preferences and will prove useful in subsequent analyses.

The CES expenditure function (2.2) can then be transformed into an individual-level econometric specification for each subject s and each modified dictator game i :

$$p_{s,d}^i \pi_{s,d}^i = \frac{\gamma_s}{(p_{s,d/r}^i)^{\kappa_s} + \gamma_s} + \epsilon_s^i \quad (2.3)$$

with the error term⁴¹ ϵ_s^i assumed to be (independent and identically) normally distributed with zero mean and variance σ_s^2 .

We generate estimates of parameters γ_s and κ_s for each subject separately using maximum likelihood with a two-limit Tobit specification.⁴² Specifically, we implement the two-stages procedure of Fisman et al. (2007): first we estimate (2.3) by non-linear-least squares and second we use the estimated parameters from this first stage as initial values for the Nelder-Mead direct search algorithm which solves the maximum likelihood problem. Such estimates are then used to recover the estimated values of the underlying CES parameters α_s and ρ_s .⁴³

⁴¹For perfectly (homothetically-)consistent subjects, the error term can only reflect misspecification of the functional form of the utility function as, abiding by our hypotheses, the CES functional form should support their behavior. For less consistent subjects, the error term captures small deviations from the optimal budget share due to inattention, miss-clicks, computation mistakes, or other errors they make.

⁴²Corner solutions at $p_{s,d}^i \pi_{s,d}^i = 0$ and at $p_{s,d}^i \pi_{s,d}^i = 1$ are indeed theoretically grounded: they obviously correspond to the behavior of a purely selfish/altruistic subject and they also correspond to the behavior of a purely efficiency-oriented subject who will alternatively allocate everything to either the dictator or the recipient (depending on the relative price). Our data is thus subject to double "censoring" due to the likelihood of data concentration on the extrema.

⁴³The estimation process is more extensively described in appendix B.2.1.

Finally, two comments must be made regarding our estimation strategy.

First, the parameter identifying the equality-efficiency tradeoff ρ_s cannot be credibly estimated when the selfishness parameter α_s is either close to 1 (pure selfishness) or to 0 (pure altruism). In order to cope with this problem, we stick to Fisman and coauthors' strategy of excluding subjects with an average endowment share kept/given above a specific threshold. We adopt the 99% threshold the authors mention in their most recent studies (Fisman et al., 2015a,b, 2017) and set these subjects' parameter to either $\hat{\alpha}_s = 1$ (purely selfish subjects) or $\hat{\alpha}_s = 0$ (purely altruistic subjects).⁴⁴

Second, as mentioned above, we impose further restrictions on the estimation to take into account violations of the underlying utility model. We consider two restriction strategies: either we run the estimation on subjects whose overall behavior can be deemed close enough to optimizing a homothetic utility function ("subjects-based restrictions"), or we run the estimation on the whole sample but only on the subset of observations that do not significantly violate the revealed preference axioms for each subject ("observations-based restrictions"). In the first case, we retain either $HEI \geq 0.51$ ⁴⁵ or average efficiency level in the HEV ≥ 0.927 ⁴⁶ as the threshold for excluding subjects falling too far away from homothetic preferences.⁴⁷ In the second case, we retain either all observations not violating HARP or observations associated with an efficiency level ≥ 0.95 in the HEV for each subject.⁴⁸ Most of the

⁴⁴The 99% threshold amounts to excluding a total of 32 subjects (9.07%), among which 30 "purely selfish" and 2 "purely altruistic" subjects. For comparative purposes, 19.7% of subjects were excluded in the baseline treatment of in Fisman et al. (2015a) (based on Fisman et al. (2007)'s experiment), which is comparable to our 17.9% of excluded Students. The alternative 95% threshold used in the original 2007 experiment amounts to excluding 25.2% of our student sample (13.6% in the full sample), versus 26.3% in Fisman and coauthors' sample.

⁴⁵Such a restriction amounts to excluding 106 out of the 353 subjects (30.03%: 54 Politicians (38.85%), 20 Students (16.26%) and 32 Non-Students (35.16%).

⁴⁶Such a restriction amounts to excluding 66 out of the 353 subjects (18.7%: 34 Politicians (24.46%), 9 Students (7.32%) and 23 Non-Students (25.27%).

⁴⁷We select the two thresholds based on power computations which are detailed in appendix B.1.3. We rely on the predictive success measure (Beatty and Crawford, 2011), which is defined as the difference between an indicator of whether subject s passes the HARP test at homothetic efficiency level h_s^i and the proportion of a hundred simulated random players who pass the same test. We compute the predictive success for each subject and for a variety of h_s^i to find the level that maximizes the predictive success on the entire sample.

⁴⁸Such a choice means only keeping observations individually responsible for a 5% efficiency loss with respect to homothetic preferences. Obviously, the 5% threshold is an arbitrary value. Yet, resorting to power computations to find a more scientifically-grounded threshold is much more difficult with vectors than with single numbers. Such endeavour is beyond the scope of this chapter.

literature implements only the first strategy and considers removing subjects based on the Afriat's Efficiency Index (AEI Afriat, 1972, 1973). Our approach takes into account the consistency with homothetic preferences instead of merely consistency with some preferences and considers the compatibility of the results with other definitions of "closeness" to optimizing behavior, building in particular on the richness of the HEV.

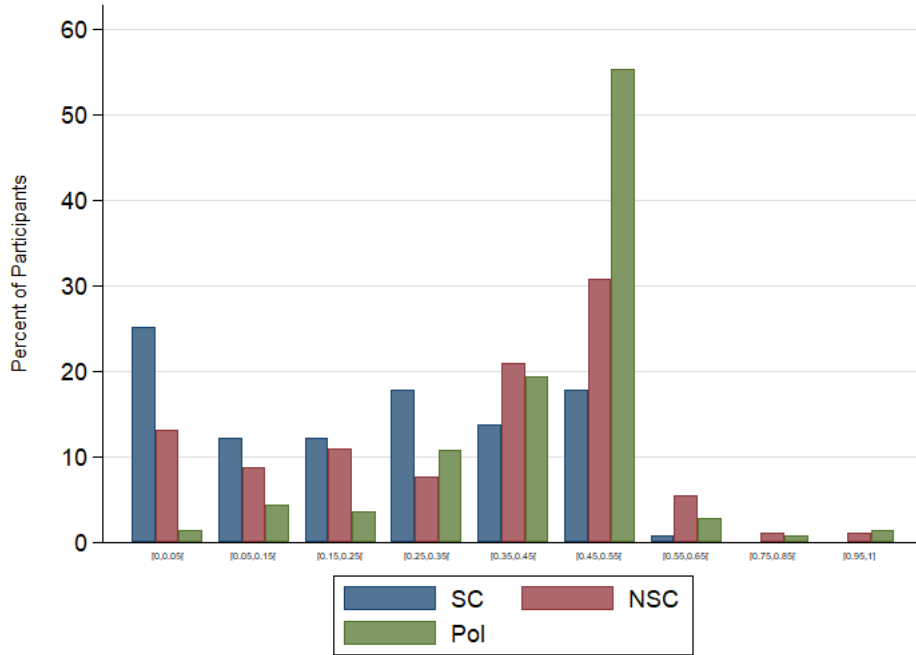
5 Results

This section exposes our results. We start with a brief overview of the data at hand without imposing much statistical structure (sub-section 5.1). Then, we turn to more elaborate analyses based on individually-estimated parameters from the structural utility model (sub-section 5.2). Finally, we provide complementary analyses on our results (sub-section 5.2.3).

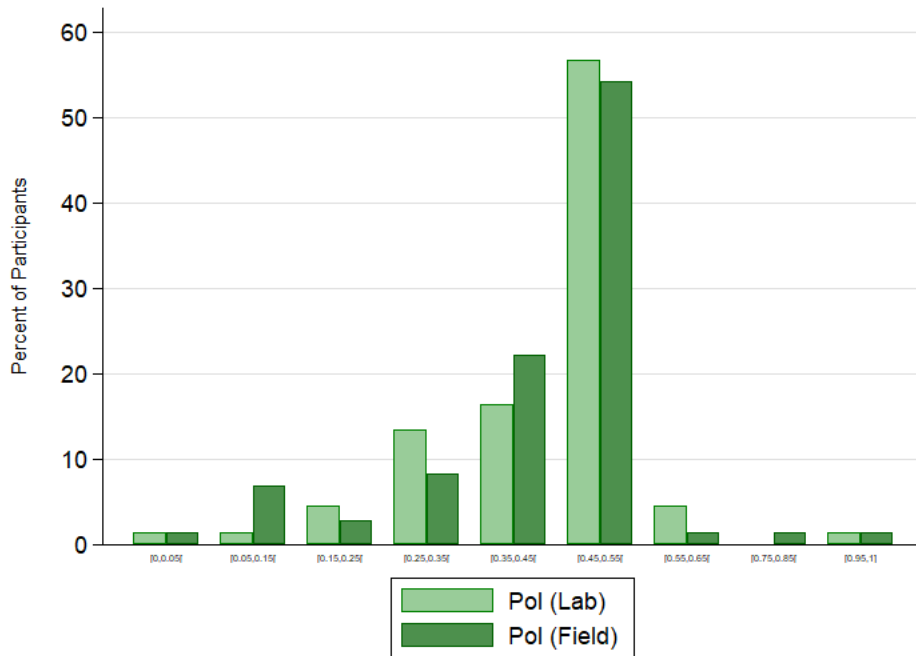
5.1 Raw data overview

We begin by a graphical overview of the data we collected without imposing any additional structure. The figure 2.2 plots the distribution of the fraction of tokens passed to the recipient $\pi_{s,o}^i / (\pi_{s,d}^i + \pi_{s,o}^i)$, averaged at the subject level.

Figure 2.2: Distribution of average pass fraction



(a) Main groups



(b) Politicians' sub-groups

Note: Pol = Politicians. SC = Student Control. NSC = Non-student Control

Most noticeably, we uncover clear behavioral differences between Politicians and both control groups (see figure 2.2a, that is panel (a) in figure 2.2). Only few Politicians (1,44%) allocate most tokens to themselves (hence a pass fraction below 5%), whereas the bulk of them target a rather equal allocation of money between themselves and the recipient (55.4%). On the contrary, both control groups appear more selfish and less fair-minded: 13.19% of the Non-Students and 25.2% of the Students allocate most tokens to themselves, whereas 30.77% Non-Students and 17.89% Students achieve approximately an equal split. The distributions of the average pass rate across groups shows similar patterns, except they are somewhat smoother. In particular, Politicians spend an average 43.24% of the endowment on the recipient, hence roughly 10 percentage points more than Non-Students (34.06%) and 20 percentage points more than Students (22,2%).⁴⁹

All statistical tests are highly significant for all pairwise group comparisons (two-sided t-test, Fligner-Policello test and Kolmogorov-Smirnov test: $p < 0.01$).⁵⁰ Moreover, we do not find any statistical difference between Politicians from the Lab and Politicians from the Field (see figure 2.2b): the latter give somewhat less than the former but the statistical tests all fail to detect such difference.⁵¹ Overall, Politicians thus display a higher concern about others than each of the control groups. Interestingly, there is approximately the same proportion of subjects favoring the recipient over themselves on average between Politicians and Non-Students (8.63%

⁴⁹The result for Students is fairly close to the 20% norm from standard split-the-pie dictator games with convenience samples (Camerer, 2011, Engel, 2011). As an additional point of comparison, Fisman et al. (2007) report an average pass rate of about 21%. We take these comparisons as first piece of evidence that our experiment yields quantitatively similar results to previous findings with similar populations. Furthermore, comparing Students and Non-students provides comparable patterns with respect to the literature: students are more likely to give nothing, less likely to give everything, and less likely to achieve equal payoffs (see the meta-analysis on dictator games by Engel, 2011).

⁵⁰Abiding by the same logic as in chapter 1, we report three main statistical tests to assess the statistical significance of the differences between our groups. Two tests on central tendencies: the standard parametric t-test (mean differences), the non-parametric robust rank-order test from Fligner and Policello (1981) (median differences). And one test on differences in entire distributions: the non-parametric Kolmogorov-Smirnov test. As a robustness check, applying the commonly-employed Mann-Whitney test on medians does not change any of our results.

⁵¹Pass fraction (pass rate): t-test: $p = 0.648$ (0.797) ; FP test: $p = 0.451$ (0.91) ; KS test: $p = 0.484$ (0.756). We also conduct Monte-Carlo simulations to assess the power of our statistical tests for comparing the Politicians' sub-groups. As shown in appendix B.3.2, we can convincingly rule out any effect greater than 7.5 percentage points and arguably uncover a 5 percentage points difference (although the odd is lower than 90%). Our tests thus seem rather powerful in detecting a differentiated behavioral pattern between Politicians from the Lab and Politicians from the Field.

and 6.59% with pass rate ≥ 0.55 , respectively). In other words, observed differences are not particularly driven by Politicians giving away money⁵² but rather by Politicians acting less self-oriented and more fair-minded overall.⁵³ As such, these aggregate results already pinpoint that motivations out of pure private self-interest (or "intrinsic motivations") seem to matter more to politicians than to other people.

However, these distributions do not provide much information as to the way our groups trade off equality against efficiency. As a consequence, we further look at the relationship between the average hold rate and the relative price over all decisions made within each group. Indeed, an equality-oriented subject will react in an opposite way to an increase in the relative price of keeping the money as compared to an efficiency-oriented subject (see sub-section 3.2). We find that a 1 unit increase in the relative price is related to a 0.02 unit increase in the hold rate among Politicians (Pearson's correlation coefficient). In other words, Politicians overall appear to slightly favor the role whose tokens are the most expensive (i.e., the dictator when the relative price is high and the recipient when it is low), thus favor payoffs' equality. Yet, we cannot rule out that such coefficient differs from zero ($p = 0.291$). On the contrary, we find a negative and significantly-different-from-zero correlation ($p < 0.01$) among both Non-Students (-0.06) and Students (-0.11), which suggests a sensitivity to total payoff's maximization (efficiency). At first glance, it thus seems that Politicians are less likely to prefer efficiency over equality than both control groups. Such results at the aggregate level thus anticipate on our main findings below.

Though comparing aggregate data is useful to grasp first-hand evidence of cross-group differences, they actually tell us little about individual behavior. Indeed, previous results point towards the large degree of individual heterogeneity in distributional preferences, which may aggregate in a variety of ways entailing a likely bias

⁵²Indeed it could have been the case that Politicians shy away from keeping money had they not taken the game seriously. When added to the fact that most of our data can be rationalized by a well-behaved utility function (chapter 1), such a finding subsequently suggests that our experiment actually reached a decent level of saliency, even for Politicians.

⁵³The specific recruitment process that we implemented for both Politicians from the Lab and Non-Students may introduce some contamination across early and late sessions, which may bias our results. Indeed, in order to recruit enough subjects, we partly relied on word-to-mouth communication between past and potential subjects. We check for such possible cross-sessions contagion effect in appendix B.3.1 and find no evidence in favor of a sizable bias.

on the global picture (Andreoni and Miller, 2002, Fisman et al., 2007, 2017, Müller, 2019, Kerschbamer and Müller, 2020). We thus further classify subjects based on the compatibility of their decisions with pure distributional types. Specifically, we classify a subject selfish (resp. altruist) if the average hold (resp. pass) rate is above a given threshold (either 0.95 or 0.99, following Fisman et al., 2015a). We also classify a subject efficiency-oriented (resp. equality-oriented) if the fraction of choices consistent with efficiency-orientation (resp. equality-orientation) is above a given threshold. Specifically, we first classify choices as consistent with either efficiency or equality. We consider a choice to be consistent with efficiency if it allocates everything to the role that can earn the most (depending on the relative price). We consider a choice to be consistent with equality if it allocates the same amount to both roles. We also implement a one-ECU tolerance in order to account for mouse imprecision or small mistakes. Then, we classify subjects as efficiency-oriented (resp. equality-oriented) if 95% or 99% of their choices are classified as consistent with efficiency (resp. equality).

Table 2.1: Canonical distributional types by groups

	Threshold = 0.95				Threshold = 0.99			
	Selfishness	Altruism	Efficiency	Equality	Selfishness	Altruism	Efficiency	Equality
Student Control ($N = 123$)	31 (25.20)	0 (0.00)	0 (0.00)	1 (0.81)	22 (17.89)	0 (0.00)	0 (0.00)	1 (0.81)
Non-student Control ($N = 91$)	11 (12.09)	1 (1.10)	0 (0.00)	5 (5.49)	7 (7.69)	0 (0.00)	0 (0.00)	2 (2.20)
Politicians ($N = 139$)	3 (2.16)	2 (1.44)	1 (0.72)	17 (12.23)	1 (0.72)	2 (1.44)	1 (0.72)	11 (7.91)

Note: The numbers indicate the headcounts of each category within each group. The percentages in parentheses indicate the proportion of subjects such headcounts represent in the total of the group, including unclassified subjects. These subjects accounts for about 80% of our sample with the 0.95 threshold (83.45% of the Politicians, 81.32% of the Non-Students and 73.99% of the Students).

The distributional types are determined based on the proximity of each subject's average behavior with one strict distributional concerns. An average hold (resp. pass) rate above 0.95 or 0.99 is indicative of a selfish (resp. altruistic) subject. A behavior compatible with efficiency (resp. equality) 95% or 99% of the time is indicative of an efficiency-oriented (resp. equality-oriented) subject.

The table 2.1 displays the results of such exercise. First, in coherence with

what we previously observed, the most frequent distributional type for Politicians is equality-orientation, whereas it is selfishness for Students. Non-Students strike a balance between the two groups, but with a slightly bigger tendency towards selfishness. Interestingly, there is approximately the same proportion of selfish Non-Students than the proportion of equality-oriented Politicians: about 7% for the 0.99 threshold and about 12% for the 0.95 threshold. Second, we are overall able to classify only up to 20.4% subjects with the least-demanding threshold. Indeed, such a classification omits that subjects may have several distributional preferences in mind. For instance they can be willing to sacrifice part of their payoffs for the sake of equality or efficiency, but without strictly abiding by the pure Equality/Efficiency prediction. Finely accounting for such intermediate types of preferences requires more evolved techniques than basic descriptive statistics. The next sub-sections thus compare our groups based on the parameters we obtained from estimating a CES utility function at the individual level.

5.2 Between-groups comparisons in estimated parameters

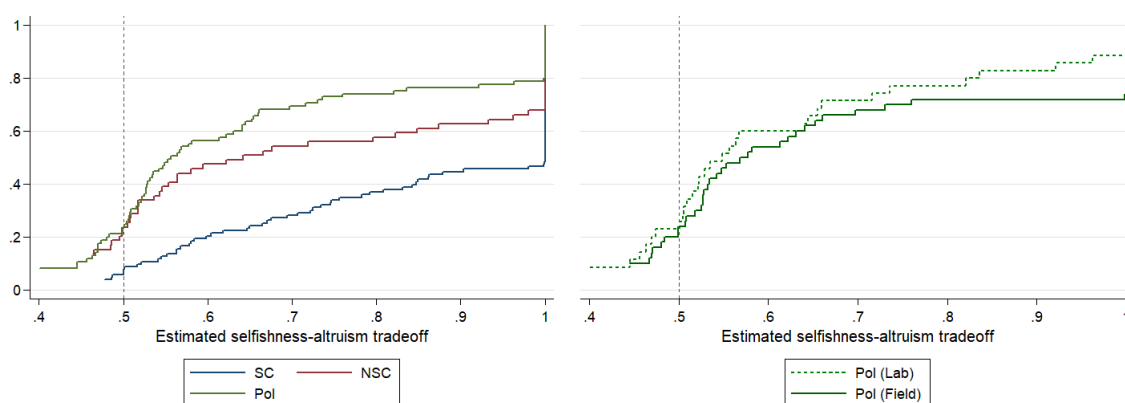
Skimming through the data by imposing very little structure already indicates quite different behavioral patterns across our groups. However, such exercise cannot fully take into account individual heterogeneity nor reveal which differences in underlying distributional preferences may be at play. Indeed, apart in the very specific context where only one distributional concern matters to subjects, which is not the case for 80% of our sample, we cannot so far accurately identify the behavioral motives. In order to shed more light on differences in distributional preferences between our groups, we estimate the structural utility model exposed in section 4.2 to disentangle two preference parameters: a first parameter ($\hat{\alpha}_s$) indicates how much weight subject s places on both himself and the recipient, while a second parameter ($\hat{\rho}_s$) indicates how subject s trades off reducing payoff differences (equality) and increasing total payoff (efficiency). Furthermore, the analysis in sub-section 5.1 does not account for choice inconsistencies. Since ignoring these inconsistencies may introduce some noise in the cross-groups comparisons, we screen for inconsistent subjects or decisions prior to running the structural estimation. To do so, we rely on the non-parametric revealed preference techniques from section 4.1 to define several restriction strategies. In the following, we first study the parameter identifying the

Selfishness/Altruism tradeoff ($\hat{\alpha}_s$), and second we examine the parameter identifying the Equality/Efficiency tradeoff ($\hat{\rho}_s$).

5.2.1 Selfishness/Altruism tradeoff

The figure 2.3 plots the cumulative distribution functions of estimated parameter $\hat{\alpha}_s$ by groups for subjects with $\text{HEI} \geq 0.51$.⁵⁴

Figure 2.3: CDF of estimated Selfishness/Altruism parameter ($\hat{\alpha}_s$) by groups ($\text{HEI} \geq 0.51$)



(a) Main groups

(b) Politicians' sub-groups

Note: 16 out of 247 subjects have $\hat{\alpha}_s \leq 0.4$: 6 Politicians (2 from the Lab, 4 from the Field), 7 Non-students and 3 Students.

The dotted line identifies $\hat{\alpha}_s = 0.5$, that is the value for which the subject equally weighs himself and the recipient.

First, we find Politicians to weigh more themselves than the recipient overall

⁵⁴The Homothetic Efficiency Index (HEI) quantifies the amount by which all budget lines must be shifted in order for the subject to be consistent with homothetic preferences. Based on a predictive success measure of the homothetic utility maximization model for our data, we restrict the analysis to subjects whose HEI is above 0.51. We also implement other restrictions based on alternative measures of closeness to homothetic utility maximization. For the sake of parcimony, we graphically display only the results based on the HEI restriction. The cumulative distribution functions of the estimated parameters for the other restrictions on the data can be found in the appendix B.4.1. Although the absolute values of the parameters fluctuate with the restriction strategy under scrutiny, the relative position of a group's CDF with respect to the other groups is similar across strategies for both parameters.

(average estimated parameter for Selfishness/Altruism of 0.633), which is somewhat consistent with the usual self-interest assumption in economic theory. We further classify subjects into 4 categories based on values of $\hat{\alpha}_s$ following Fisman et al. (2015a, 2017): near-selfishness ($\hat{\alpha}_s \geq 0.95$), intermediate selfish-orientation ($\hat{\alpha}_s \in (0.55, 0.95)$), fair-mindedness ($\hat{\alpha}_s \in (0.45, 0.55)$) and altruistic-orientation ($\hat{\alpha}_s \leq 0.45$). Approximately a quarter of Politicians appear near-selfish (about 22% with $\hat{\alpha}_s > 0.95$). However, only a slim majority of the Politicians shows a true tendency to favor oneself over the recipient (about 51% with $\hat{\alpha}_s > 0.55$).

We observe some differences between between Politicians from the Lab (average $\hat{\alpha}_s$ of 0.616) and Politicians from the Field (average $\hat{\alpha}_s$ of 0.645) in the figure 2.3b. In particular, the two distributions are close to one another until the parameter gets close to 1. Indeed, the proportion of selfish types with $\hat{\alpha}_s \geq 0.95$ is higher among Politicians from the Field (28%) than among Politicians from the Lab (14%). We uncover similar results when scrutinizing the distribution functions obtained with alternative restriction strategies (see figure B.10 in appendix B.4.1).

Nonetheless, these differences between the two Politicians' sub-samples are generally not statistically significant at conventional levels (right part of table 2.2). There is one exception though: when implementing the restriction based on the average HEV, the proportion of selfish types do significantly differ. We conclude from these observations that the two Politicians' sub-samples are globally alike, though the subjects that came into the laboratory were less strictly egoistic.⁵⁵

Result 1 (selfishness v. altruism among Politicians' sub-groups): *Self-interest dominates fair-mindedness and altruism among Politicians, but only marginally. Field and Lab Politicians similarly trade off selfishness and altruism, except for very high levels of selfishness.*

Second, the distribution for Politicians is almost systematically above the distribution for both our control groups (figure 2.3a). This indicates an overall concern for others among Politicians greater than among both Non-students (average $\hat{\alpha}_s$ of 0.676) and Students (average $\hat{\alpha}_s$ of 0.828). In other words, both the average

⁵⁵The latter result is interesting insofar as it indicates that Politicians from the Field, which were more likely to play with an acquaintance due to the recruitment process, did not actually weigh their partner more than Politicians from the Lab. Reputation mechanisms or other peer effects could indeed have triggered an increase in altruistic behavior in the field which we do not observe.

Politician and the average Non-student value the recipient as approximately half of herself, whereas the average Student values the recipient as a fourth of herself. In addition, Politicians appear even farther away than Non-Students from Students, who are predominantly selfish.

Furthermore, eye-balling suggests that the differences in the empirical distributions are essentially driven by two elements. First, there is a lower concentration of Politicians at higher percentiles with a sharper jump near $\hat{\alpha}_s = 1$ among both control groups. Indeed there are 19 near-selfish Politicians (23.38%) against 21 Non-students (35.59%) and 56 Students (54.37%). Second, the proportion of fair-minded subjects is higher among Politicians than among both control groups: 32 Politicians (37.65%) against 16 Non-students (27.12%) and 10 Students (9.71%). Interestingly, there is approximately the same proportion of near-selfish Non-Students than of fair-minded Politicians, and conversely. On the contrary, the proportions of rather-altruistic and self-oriented subjects are similar across the two groups: there are 9 Politicians (10.59%) with $\hat{\alpha}_s < 0.45$ against 7 Non-Students (11.86%) and 25 Politicians (29.41%) with $\hat{\alpha}_s \in (0.55, 0.95)$ against 15 Non-students (25.42%). Consequently, the group differences are essentially driven by Politicians being less selfish and more fair-minded, which is consistent with what we observed from raw data.

Table 2.2: p -values from statistical tests for differences in Selfishness/Altruism trade-off $\hat{\alpha}_s$

		Pol v. NSC				Pol v. SC				Lab v. Field			
		HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v
$\hat{\alpha}_s$	t-test	.362	.377	<0.01	.026	<0.01	<0.01	<0.01	<0.01	.605	.256	.623	.916
	FP test	.223	.187	.016	.017	<0.01	<0.01	<0.01	<0.01	.396	.256	.927	.734
	KS test	.2	.162	.023	.035	<0.01	<0.01	<0.01	<0.01	.559	.162	.975	.873
Prop $\hat{\alpha}_s \geq 0.95$	Test of proportions	.081	.113	<0.01	.023	<0.01	<0.01	<0.01	<0.01	.135	.015	.397	.31
	F-test	.091	.12	<0.01	.033	<0.01	<0.01	<0.01	<0.01	.187	.02	.455	.41
Prop $\hat{\alpha}_s \in (0.45, 0.55)$	Test of proportions	.187	.123	.202	.051	<0.01	<0.01	<0.01	<0.01	.708	.914	.618	.735
	F-test	.212	.135	.253	.064	<0.01	<0.01	<0.01	<0.01	.821	1.000	.725	.862
Prop $\hat{\alpha}_s < 0.45$	Test of proportions	.811	.791	.46	.964	.032	.018	<0.01	.012	.833	.58	.642	.887
	F-test	.795	.807	.586	1.000	.039	.025	<0.01	.013	1.000	.751	.666	1.000
Observations		144	173	230	230	188	219	262	262	85	105	139	139

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$. t-test: two-samples Student test. FP: Fligner-Policello. KS: Kolmogorov-Smirnov. F-test: Fisher's exact test.

However the statistical tests yield mixed results (p-values in the HEI columns from table 2.2). On the one hand, all differences with respect to Students are highly statistically significant: tests on central tendencies, entire distributions or proportions all reject the null hypothesis of no difference between Politicians and Students (middle part of table 2.2). On the other hand, almost all tests fail to reject the null when comparing Politicians with Non-Students, except for testing for equal proportions of near-selfish subjects (left part of table 2.2).

Yet, when comparing groups based on alternative restriction strategies, we do observe significant differences between Politicians and Non-Students. In particular, we observe a difference between subjects-based and observations-based restrictions. When implementing a restriction strategy that excludes inconsistent observations instead of inconsistent subjects, almost all the tests confirm our observations (p-values in the H²MI and HEV^v columns from table 2.2): differences in central tendencies, overall distributions and proportion of near-selfish subjects all are significant, and the difference in the proportion of fair-minded subjects is significant with the HEV^v restriction. At the same time, the relative differences between the groups remain essentially unchanged (see figure B.9 in appendix B.4.1) and the tests on Politicians v. Students comparisons still yield highly significant differences.

One explanation for such a discrepancy in the results of the statistical tests when different restriction strategies are used is the power loss involved by removing subjects falling too far away from homothetic utility maximization. On the contrary, observations-based restrictions maintain all subjects in the dataset. In order to shed more light on that issue, we implement the same tests to the whole sample without any limitation to homothetic-enough data. The overall picture is unchanged with respect to the distribution of the parameters across groups (figure B.13 in appendix B.4.1). Besides, our results are consistent with little significant differences in central tendencies (either mean or median), which may hide significant differences in entire distributions in the above-mentioned way: Politicians are less likely to endorse a

purely selfish behavior and more likely to balance both roles equally (table B.5 in appendix B.4.1).⁵⁶

Considering this whole bunch of observations, we draw the following conclusions:

Result 2 (selfishness v. altruism in Politicians v. Non-Students): *Politicians appear less self-oriented than a comparable control group. The evidence is mixed regarding central tendencies, but Politicians tend to favor fair-mindedness ($\hat{\alpha}_s \in (0.45, 0.55)$) over pure selfishness ($\hat{\alpha}_s \geq 0.95$) more often than Non-Students.*

Result 3 (selfishness v. altruism in Politicians v. Students): *Politicians are significantly less self-oriented than a student control group.*

With respect to our working hypotheses exposed in section 2, these results lead to accept H^1 as Politicians do not systematically implement the self-interested option. But they do so in higher proportions than we expected. Indeed, our results also lead to a rejection of H^2 : Politicians are at least less strictly selfish than Non-Students, if not more fair-minded. In other words the two groups do differ in some ways. Besides, H^3 is clearly accepted as Politicians are very significantly more altruistic than Students.

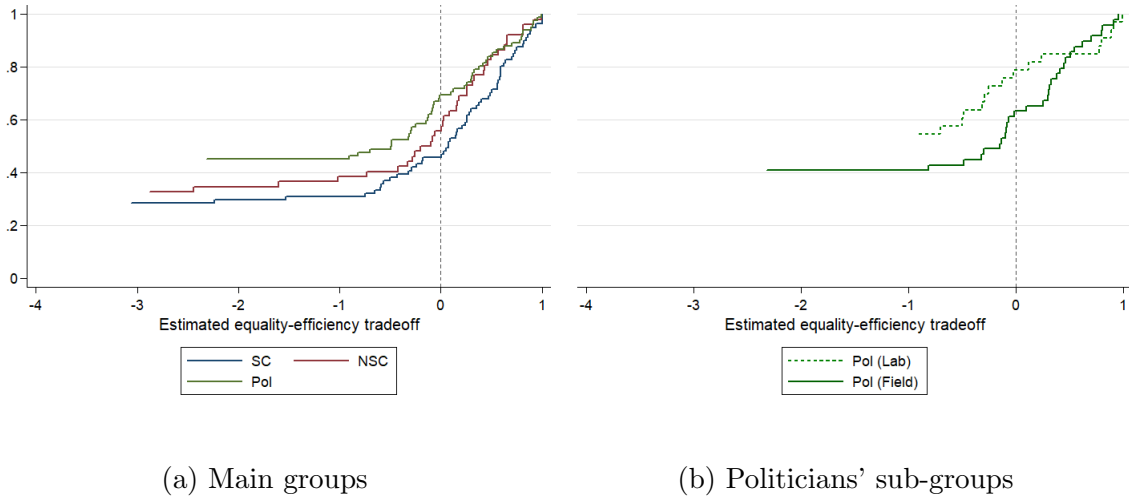
5.2.2 Equality/Efficiency tradeoff

The figure 2.4 plots the cumulative distribution functions of estimated parameter $\hat{\rho}_s$ by groups for subjects with $HEI \geq 0.51$.

First, Politicians display a higher concern for equality as compared to efficiency (median estimated parameter for Equality/Efficiency of -0.86). More specifically, taking 0 as the theory-driven threshold for disentangling equality-oriented prefer-

⁵⁶Indeed, the KS test rejects the null that the two distributions are the same at the 5% significance level, similar to what we find with observations-based restrictions. The tests on proportions of near-selfish and fair-minded subjects also yield significant differences. We also recover the corresponding results when comparing Politicians' sub-groups: all tests are not significant except for tests on differences in entire distributions or in the proportion of selfish subjects. Noticeably, the magnitude of the differences as well as the significance level are higher when comparing Politicians to Non-Students than when comparing Politicians' sub-groups, which we consider as supplementary evidence that Politicians and Non-Students differ in how they trade-off selfishness and altruism.

Figure 2.4: CDF of estimated Equality/Efficiency parameter ($\hat{\rho}_s$) by groups (HEI ≥ 0.51)



Note: 74 out of 215 subjects have $\hat{\rho}_s \leq -4$: 36 Politicians (17 from the Lab, 19 from the Field), 16 Non-students and 22 Students. The dotted line identifies $\hat{\rho}_s = 0.5$, that is the value which separates equality-orientation (< 0) from efficiency-orientation (> 0).

ences ($\hat{\rho}_s < 0$) from efficiency-oriented preferences ($\hat{\rho}_s > 0$), we uncover that approximately two-thirds Politicians (69.51%) are equality-oriented. However, we observe some differences between the two Politicians' sub-groups as the distribution of $\hat{\rho}_s$ from the lab data (median $\hat{\rho}_s$ of -0.996) is systematically above the one from the field data (median $\hat{\rho}_s$ of -0.652) until $\hat{\rho}_s \approx 0.5$ (figure 2.4b). As a consequence, Politicians from the Lab lean more towards equality considerations than Politicians from the Field. Furthermore, there are 26 equality-oriented Politicians from the Lab (78.79%) against 31 Politicians from the Field (63.27%). We uncover similar results when scrutinizing the distribution functions obtained with alternative restriction strategies, except the gap between the two functions is systematically smaller (see figure B.12 and B.13 in appendix B.4.1).

However, only the test on differences in means yield p-values lower than 10% (first line in the right part of table 2.3). Considering that the distributions of $\hat{\rho}_s$ are highly

Table 2.3: p -values from statistical tests for differences in Equality/Efficiency trade-off $\hat{\rho}_s$

		Pol v. NSC				Pol v. SC				Lab v. Field			
		HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v
$\hat{\rho}_s$	t-test	.242	.517	.66	.734	.202	.281	<0.01	.023	.023	.017	.463	.408
	FP test	.604	.295	.044	.154	.081	.053	<0.01	<0.01	.421	.963	.79	.205
	KS test	.585	.134	.152	.15	.02	.019	<0.01	<0.01	.216	.334	.824	.231
Prop $\hat{\rho}_s < 0$	Test of proportions	.106	.023	.05	.059	<0.01	<0.01	<0.01	<0.01	.134	.455	.472	.483
	F-test	.139	.03	.068	.066	<0.01	<0.01	<0.01	<0.01	.151	.524	.59	.574
Observations		134	163	192	182	163	194	211	190	82	102	124	115

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$. t-test: two-samples Student test. FP: Fligner-Policello. KS: Kolmogorov-Smirnov. F-test: Fisher's exact test.

skewed with a long tail on the left, the mean is a relatively uninformative statistic.⁵⁷ Furthermore, we obtain at least the same statistical results when implementing other restriction strategies, if not totally insignificant results. As a consequence, we do not take these differences as conclusive evidence that the two sub-groups differ with respect to the Equality/Efficiency tradeoff.

Result 4 (equality v. efficiency among Politicians' sub-groups): *Politicians show an overall tendency to favor equality over efficiency. Field and Lab Politicians similarly trades off both distributional concerns.*

Second, the distribution for Politicians is almost systematically above the distribution for both our control groups, with a rather consistent ordering of the three distributions throughout (figure 2.4a). This indicates a greater overall concern for equality over efficiency among Politicians than among both Non-students (median $\hat{\rho}_s$ of -0.312) and Students (median $\hat{\rho}_s$ of 0.005). Indeed, there are 57 Politicians with equality-oriented preferences (69.51%) against 29 Non-Students (55.77%) and 37 Students (45.68%). In other words, Students slightly favor efficiency over equal-

⁵⁷Relatedly, we observe stark differences between Politicians and Students below but the differences in means are sometimes not significant, depending on the underlying restriction strategy. We thus do not consider the t-tests to provide compelling information about the statistical significance of observed differences in $\hat{\rho}_s$ but still keep them for the sake of completeness.

ity ($\hat{\rho}_s > 0$) whereas the contrary holds for Non-Students and only about one third of the Politicians support efficiency.

The statistical non-parametric tests on the distribution of $\hat{\rho}_s$ confirm our results on Politicians v. Students (middle part of table 2.3). Yet, comparing Politicians with Non-Students leads to failing to reject the null in all cases but one. (left part of table 2.3) At the same time, the reduced-form test based on the position of the parameter with respect to 0 provides a much clearer picture. Politicians and Students systematically differ very significantly no matter the restriction strategy nor the statistical test under scrutiny (test of proportions or Fisher's exact test). Concerning Politicians and Non-Students, the statistical evidence is weaker but we still observe a significant difference for three out of four restriction strategies. Only the subject-based restriction on the HEI yields a non-significant result at the 10% level.

Again, such discrepancy may be explained by the loss of subjects implied by subjects-based restrictions. Indeed, the significance of the tests varies widely when comparing subjects-based restrictions with observations-based restrictions. Considering the full sample without any restriction does indeed provide consistent results with observations-based restriction: Politicians are significantly more equality-oriented ($\hat{\rho}_s < 0$) than Non-Students at the 10% significance level (see table B.5 in appendix B.4.1). At the same time, the overall picture on the distribution of $\hat{\rho}_s$ is unchanged (see figure B.13) and we still detect a highly-significant difference between Politicians and Students.

We summarize these observations on parameter $\hat{\rho}_s$ between our main experimental groups as:

Result 5 (equality v. efficiency in Politicians v. Non-Students): *Politicians are more equality-oriented than a comparable control group.*

Result 6 (equality v. efficiency in Politicians v. Students): *Politicians are very significantly more equality-oriented than a student control group.*

With respect to our working hypotheses, we clearly reject all of them. Indeed, we find opposite results with respect to what we expected based on the existing literature: Politicians are not favoring efficiency over equality but the converse (\mathbf{H}^4),

and are significantly more sensitive to equality than both control groups (H^5 and H^6).

5.2.3 Additional analyses and robustness checks on estimated parameters

This section provides complementary analyses on our results. First we check our results' robustness with respect to existing results in the literature. Second, we control for a possible differentiated performance of the underlying utility model in predicting our groups' behavior.

It is noteworthy that the distribution of both estimated parameters covers a wide range of values: for the Selfishness/Altruism tradeoff it covers the whole $[0, 1]$ interval while for the Equality/Efficiency tradeoff it reaches highly negative values and gets really close to its upper limit of 1. In other words we do observe a great deal of individual heterogeneity in both parameters, which is consistent with previous studies. We also look at the within-subject correlation between estimated parameters to determine whether the way we conceptualize observed behavior into two different tradeoffs is empirically sound. To take into account the importance of outliers in the distribution of parameters, we classify subjects into 4 categories of Selfishness/Altruism (near-selfishness, intermediate selfishness, fair-mindedness and altruism) and 2 categories of Equality/Efficiency (either equality-oriented or efficiency-oriented) as defined by Fisman et al. (2015b, 2017). The classification is detailed in the appendix B.4.1 and the figure B.4.2 summarizes the data for each restriction strategy. Though we strongly reject the independence assumption between the two sets of categories (all Fisher's exact test yields $p < 0.01$), we do not observe a straightforward relationship. In particular, we find substantial heterogeneity in individual profiles with both equality-oriented and efficiency-oriented subjects being identified in all categories of Selfishness/Altruism. The two tradeoffs thus seem to correspond to empirically different patterns of choice.

In addition, we compare the estimated parameters we obtain from Students with the estimated parameters reported in Fisman et al. (2007)'s appendices. We plot the cumulative distribution functions of the parameters in the figure B.15 in appendix B.4.3. We find a higher proportion of near-selfish subjects as well as lower

levels of the Equality/Efficiency tradeoff in our student sample than in original sample. However, we do not find a statistically-significant difference with respect to either the Selfishness/Altruism tradeoff ($\hat{\alpha}_s$) or the probability to classify subjects as efficiency-oriented ($\hat{\rho}_s > 0$).

We conclude from such comparison that our experiment yields similar results to existing ones based on the standard laboratory population (students). This is particularly valuable in light of the specific payment scheme we used. Paying only a third of subjects per session bears the risk of diluting the incentives, especially when combined with paying only one decision. Subjects could thus have adopted a less self-serving behavior and favor what they think is a moral behavior to signal themselves. As emphasized by Stahl and Haruvy (2006), the benefits of moral signaling are unchanged with respect to a pay all baseline but the costs are reduced: it creates a free-lunch opportunity for moral behavior. Yet this is not what we find: our student sample comes close to what was previously observed and, if anything, it appears slightly more rather than less selfish. Our experiment thus provides additional evidence in favor of using alternative payment schemes instead of the usual-but-costly pay-all scheme in distributional studies.

Finally, we check whether our group comparisons may be confounded by different error distributions from the estimation process. For each observation, we compute the prediction error as the difference between the observed hold rate and the predicted value of the hold rate calculated with the CES estimated parameters. We observe that the distributions of prediction errors are normal-shaped with a mode around zero for all restrictions. Our model thus seems to fit the data well, no matter the implemented restriction on homothetic preferences, and the empirical distributions match our theoretical assumption about the error term. Furthermore, we do not find much difference in prediction error averaged at the individual level between Politicians and Non-Students. If anything, our model under-estimate the hold rate for Politicians relative to Non-Students, which ultimately means that the observed difference between the groups is likely conservative. As a consequence, our results are unlikely driven by important cross-groups difference in the model's predictive accuracy. However, we also find some evidence that the prediction errors are closer to zero for Students than for Politicians as well as for Politicians from the Lab than for Politicians from the Field. The former result is rather intuitive considering that Students are closer to homothetic preferences than Politicians. We cannot think of

any particular reason explaining the latter result, but we note that it may explain why we observe some (generally not significant) differences in estimated parameters between the Politicians' sub-groups.⁵⁸ Detailed results are presented in the appendix B.4.4.

5.3 Regression framework

We now turn to an econometric analysis of the differences in estimated parameters between our groups, to assess the validity of our results when adding control variables. In particular, we find that Politicians are more educated and earn more than Non-Students, that Politicians differ from Students in every available observable, and that Politicians from the Field are less educated and more often have a fundamental science background than Politicians from the Lab.⁵⁹ These differences in observables between our groups may bias our above results based on simple statistical tests, hence the need for a complementary regression analysis. Tables 2.4 (Pol v. NSC), 2.5 (Pol v. NSC) and 2.6 (Politicians' sub-groups) present the results of regressing a group variable (dummy) on our estimated parameters, along with control variables.⁶⁰

We run a tobit specification for the Selfishness/Altruism tradeoff parameter $\hat{\alpha}_s$ and a probit specification for the Equality/Efficiency parameter $\hat{\rho}_s$ (in line with the procedures proposed by Fisman et al., 2015a,b, 2017). The former takes into account the concentration of data at $\hat{\alpha}_s = 1$ (strict selfishness), which is mostly driven by the exclusion restriction on subjects whose average behavior is close to zero giving (average share of endowment kept greater than 0.99). The latter takes into account the skewness of the parameter's distribution, which would bias OLS estimates. In

⁵⁸In addition, we compare Non-Students and Politicians from the Lab in appendix B.4.4 and globally recover the same results as with the pooled group of Politicians.

⁵⁹For a definition of our control variables and a discussed presentation of observed differences across groups, we refer the interested reader to the corresponding material in both the chapter 1 and the appendix A. As a reminder, we find that Politicians are close to Non-Students except they hold a higher degree and earn more, whereas Students differ from Politicians in all available observables (age, gender, education and major). We also observe that the two Politicians' sub-groups are close except Politicians from the Field hold a lower degree which is more often a degree in Fundamental Science.

⁶⁰For the sake of parcimony, we do not comment upon the results obtained from the control variables, which are not of primary interest. We discuss such results and document their link with the literature in appendix B.5.3.

such a case, the regression model builds on the theoretical prediction that $\hat{\rho}_s > 0$ implies that a higher share of the endowment is given to the role which can earn the most from the decision (hence maximizing total payoffs: efficiency concern), whereas $\hat{\rho}_s < 0$ implies that a higher share of the endowment is given to the role which can earn the less from the decision (hence reducing payoff differences: equality concern). In other words, we run regressions on a binary variable taking value 1 if the subject is efficiency-oriented ($\hat{\rho}_s > 0$) and 0 if she is equality-oriented ($\hat{\rho}_s < 0$). We run identical regressions for the parameters obtained based on each of our four restriction strategies.

For each parameter and each restriction strategy, we display at least two models for pairwise comparisons across groups. The first model includes only the group variable (taking value 1 for Politicians in tables 2.4 or 2.5 and for Politicians from the Field in table 2.6), whereas the second model includes "relevant" control variables.⁶¹ We consider "relevant" the variables for which we observe a statistically significant difference across our groups, namely education level and income (Pol v. NSC) or education level and field of studies (Pol v. SC and Politicians from the Lab v. Politicians from the Field). We also systematically account for gender and age (when available), in view of their empirical importance as explanatory variables in distribution games. In each table, the first model for each restriction strategy is exposed in the first column, while the second model is exposed in the second column. In addition, the table 2.4 contains a third column for each restriction strategy that adds political preferences along with the "relevant" control variables in the regression.⁶²

Globally, we obtain similar results as the ones uncovered in the previous section:

⁶¹One subject (a Non-Student) had to leave the laboratory before completing the questionnaire. As we do not have any value for his observables, we exclude such subject from the regression analysis.

⁶²Political preferences here correspond to a self-declared assessment on a 11-points Likert scale ranging from far-left to far-right. Politicians from the Field appear slightly more left-leaning than Non-Students (average placement of 4.21 against 4.31 with 0 indicating far-left preferences and 10 indicating far-right preferences), but the difference is not statistically significant (t-test: $p = 0.748$). For the sake of comparison, and though it corresponds to the 2020 election and is restricted to mayors, the average is 4.73 for the very same variable in the national survey by Foucault and Bono (2020).

the differences across groups are not importantly driven by common observables.⁶³ If anything, we note that the introduction of control variables tend to strengthen our results.⁶⁴

Similar to what we observed above, the comparisons between Politicians and Non-Students partly depend on the restriction strategy under scrutiny (table 2.4). In all models, the negative value of the coefficient on the binary group variable "Pol" is consistent with Politicians being less selfish and more equality-oriented than Non-Students. Yet, we cannot reject the null hypothesis that the coefficient is null in some cases. Again, one possible explanation is that subjects-based restrictions (based on the HEI or on the average element in the HEV) impose a too heavy reduction on the data at hand as compared to observations-based restrictions (based on the H²MI or on the HEV), hence a power issue.

For the Selfishness/Altruism tradeoff ($\hat{\alpha}_s$), the coefficient is significantly different from zero at least at the 5% significance level when implementing observations-based restrictions, no matter whether control variables are added or not (table 2.4a). However, it loses statistical significance when implementing subjects-based restrictions.⁶⁶

For the Equality/Efficiency tradeoff ($\hat{\rho}_s$), the overall picture is clearer (table

⁶³The results do not change upon removing the income variable's "Do not want to answer" category.

⁶⁴As an additional robustness check, we run the very same models by changing either the definition of the exclusion restriction on selfish/altruistic-enough subjects (from $p_{s,d}^i \pi_{s,d}^i$ or $p_{s,r}^i \pi_{s,r}^i \geq 0.99$ to $p_{s,r}^i \pi_{s,r}^i$ or $p_{s,r}^i \pi_{s,r}^i \geq 0.95$) or the underlying estimation technique (Maximum Likelihood or Non-Linear Least Squares). The figures B.19 and B.21 in appendix B.5.1 graphically display the results (specification curves). Overall, we obtain very consistent results throughout. Interestingly, adopting the 0.95 threshold for the exclusion restriction (as in Fisman et al., 2007) yields more significant results on $\hat{\alpha}_s$: Politicians appear significantly less selfish than Non-Politicians even for subjects-based restrictions. We interpret such finding as indicative that the results in table 2.4 are conservative.

⁶⁵The score cluster bootstrap procedure is the counterpart of the wild cluster bootstrap procedure (Cameron et al., 2008, Cameron and Miller, 2015) when maximum likelihood estimators are used instead of standard ordinary least squares estimators. It aims at correcting the potential bias on standard errors (hence statistical tests) when there are too few clusters. Indeed, in such a situation, the residuals are downward biased (the model overfits the data) and the confidence intervals are too narrow around the estimated coefficient (which increases the probability of rejecting the null hypothesis).

⁶⁶We note that when the binary variable is introduced alone, the coefficient is significant at the 10% level, but it does not resist the cluster bootstrap correction for the small number of sessions (very last line of the bottom part of the table).

Table 2.4: Politicians v. Non-Student Control

(a) Tobit on $\hat{\alpha}_s$

	HEI			HEV ^a			H ² MI			HEV ^v		
<u>Pol</u> (Baseline: Control)	-0.057*	-0.049	-0.022	-0.055*	-0.047	-0.004	-0.109***	-0.118***	-0.134***	-0.101***	-0.110***	-0.103**
	(0.034)	(0.035)	(0.040)	(0.033)	(0.033)	(0.032)	(0.032)	(0.036)	(0.045)	(0.032)	(0.030)	(0.040)
<u>Age</u>		-0.003	-0.003		-0.003	-0.003		-0.003	-0.004		-0.001	-0.002
		(0.004)	(0.006)		(0.004)	(0.006)		(0.003)	(0.003)		(0.003)	(0.003)
<u>Gender</u> (Baseline: Male)		0.048	0.099*		0.019	0.066		-0.023	-0.008		0.008	0.034
		(0.049)	(0.055)		(0.043)	(0.044)		(0.042)	(0.051)		(0.045)	(0.052)
<u>Higher education</u> (Baseline: No)		0.010	0.047		0.013	0.055		-0.047	-0.030		-0.048	-0.056
		(0.051)	(0.059)		(0.052)	(0.068)		(0.044)	(0.063)		(0.036)	(0.047)
<u>Income</u> (Baseline: Between 2000 € and 3000 €)												
< 2000 €		-0.100	-0.082		-0.067	-0.076		-0.003	-0.016		-0.073*	-0.084
		(0.070)	(0.086)		(0.058)	(0.069)		(0.042)	(0.057)		(0.041)	(0.051)
> 3000 €		-0.035	-0.005		-0.042	-0.035		0.086	0.039		0.050	0.031
		(0.087)	(0.118)		(0.076)	(0.114)		(0.057)	(0.096)		(0.068)	(0.104)
Do not want to answer		-0.053	-0.098		-0.024	-0.067		-0.011	-0.053		-0.072	-0.095
		(0.074)	(0.089)		(0.081)	(0.080)		(0.045)	(0.068)		(0.063)	(0.084)
<u>Political preferences</u> (Self Placement)			-0.048**			-0.041*			0.009			0.006
			(0.022)			(0.021)			(0.016)			(0.014)
Observations	143	143	106	173	173	122	229	229	160	229	229	160
LR chi2	2.864	15.978	14.245	2.776	14.063	19.511	11.487	26.846	23.040	9.753	41.535	66.647
Prob > χ^2	0.091	0.025	0.076	0.096	0.050	0.012	0.001	0.000	0.003	0.002	0.000	0.000
Bootstrapped p-value	0.135	0.179	0.576	0.147	0.196	0.910	0.013	0.020	0.023	0.008	0.006	0.036

(b) Probit on $\hat{\rho}_s > 0$

	HEI			HEV ^a			H ² MI			HEV ^v		
<u>Pol</u> (Baseline: Control)	-0.387	-0.547*	-0.367	-0.465**	-0.648**	-0.528	-0.377***	-0.471***	-0.305*	-0.366**	-0.511***	-0.375*
	(0.236)	(0.290)	(0.363)	(0.219)	(0.270)	(0.337)	(0.125)	(0.131)	(0.180)	(0.179)	(0.182)	(0.219)
	[-0.145]	[-0.203]	[-0.142]	[-0.177]	[-0.245]	[-0.206]	[-0.149]	[-0.186]	[-0.117]	[-0.146]	[-0.204]	[-0.149]
<u>Age</u>		-0.012	-0.030		-0.018	-0.040**		-0.003	-0.012		-0.004	-0.006
		(0.015)	(0.021)		(0.014)	(0.020)		(0.010)	(0.014)		(0.010)	(0.013)
		[-0.005]	[-0.012]		[-0.007]	[-0.016]		[-0.001]	[-0.005]		[-0.002]	[-0.002]
<u>Gender</u> (Baseline: Male)		-0.638***	-0.575*		-0.663***	-0.468		-0.451*	-0.173		-0.336	-0.190
		(0.240)	(0.295)		(0.243)	(0.289)		(0.238)	(0.304)		(0.207)	(0.243)
		[-0.237]	[-0.223]		[-0.250]	[-0.183]		[-0.178]	[-0.066]		[-0.134]	[-0.075]
<u>Higher education</u> (Baseline: No)		0.181	0.178		0.261	0.091		0.262	0.384*		0.503**	0.598**
		(0.214)	(0.264)		(0.228)	(0.311)		(0.187)	(0.228)		(0.198)	(0.269)
		[0.067]	[0.069]		[0.099]	[0.036]		[0.103]	[0.147]		[0.201]	[0.237]
<u>Income</u> (Baseline: Between 2000 € and 3000 €)												
< 2000 €		0.457	0.175		0.482**	0.236		0.469**	0.239		0.457**	0.369
		(0.289)	(0.329)		(0.237)	(0.315)		(0.196)	(0.201)		(0.193)	(0.267)
		[0.171]	[0.069]		[0.181]	[0.092]		[0.184]	[0.091]		[0.181]	[0.146]
> 3000 €		0.136	-0.222		0.293	0.132		0.213	-0.009		0.375	0.283
		(0.334)	(0.388)		(0.318)	(0.369)		(0.211)	(0.262)		(0.258)	(0.244)
		[0.048]	[-0.083]		[0.107]	[0.051]		[0.085]	[-0.004]		[0.149]	[0.113]
Do not want to answer		0.213	0.076		0.609	0.474		0.368	0.599		0.158	0.312
		(0.438)	(0.414)		(0.446)	(0.512)		(0.381)	(0.450)		(0.439)	(0.560)
		[0.077]	[0.030]		[0.232]	[0.186]		[0.145]	[0.213]		[0.062]	[0.124]
<u>Political preferences</u> (Self Placement)			0.225***			0.175***			0.215***			0.036
			(0.065)			(0.057)			(0.053)			(0.062)
			[0.087]			[0.068]			[0.082]			[0.014]
Observations	133	133	98	163	163	114	192	192	128	182	182	125
LR chi2	2.704	11.048	21.243	4.510	17.365	43.914	9.184	23.464	83.323	4.169	19.347	17.106
Prob > χ^2	0.100	0.137	0.007	0.034	0.015	0.000	0.002	0.001	0.000	0.041	0.007	0.029
Bootstrapped p-value	0.143	0.070	0.304	0.058	0.024	0.115	0.019	0.004	0.104	0.065	0.013	0.112

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses. Marginal effects estimated at the means of covariates in brackets.

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

The "Bootstrapped p-value" report the p-value associated with the group variable when applying the score cluster bootstrap correction for the small number of clusters (Kline and Santos, 2012)⁶⁵. The bootstrap is conducted with 4999 repetitions.

Table 2.5: Politicians v. Student Control

(a) Tobit on $\hat{\alpha}_s$

	HEI		HEV ^a		H ² MI		HEV ^v	
<u>Pol</u> (Baseline: Control)	-0.136***	-0.143***	-0.126***	-0.133***	-0.142***	-0.138***	-0.132***	-0.142***
	(0.021)	(0.021)	(0.017)	(0.018)	(0.017)	(0.017)	(0.016)	(0.016)
<u>Gender</u> (Baseline: Male)		-0.046		-0.033		-0.064		-0.054
		(0.048)		(0.046)		(0.043)		(0.047)
<u>Higher education</u> (Baseline: No)		0.069		0.069		0.016		0.039
		(0.057)		(0.055)		(0.033)		(0.046)
<u>Field of Studies</u> (Baseline: Fundamental Science)								
Econ		0.020		0.006		0.153***		-0.019
		(0.076)		(0.060)		(0.054)		(0.052)
Humanities		-0.006		0.022		0.124**		-0.024
		(0.063)		(0.072)		(0.054)		(0.065)
Other		-0.094		-0.086		0.055		-0.113**
		(0.063)		(0.062)		(0.042)		(0.046)
Observations	188	188	219	219	262	262	262	262
LR chi2	43.271	151.211	51.820	111.015	67.696	99.458	67.020	108.570
Prob > χ^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Bootstrapped p-value	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000

(b) Probit on $\hat{\rho}_s > 0$

	HEI		HEV ^a		H ² MI		HEV ^v	
<u>Pol</u> (Baseline: Control)	-0.309***	-0.464***	-0.284***	-0.408***	-0.409***	-0.488***	-0.392***	-0.483***
	(0.108)	(0.111)	(0.100)	(0.108)	(0.107)	(0.113)	(0.093)	(0.097)
	[-0.121]	[-0.181]	[-0.111]	[-0.158]	[-0.154]	[-0.184]	[-0.154]	[-0.189]
<u>Gender</u> (Baseline: Male)		-0.384*		-0.447***		-0.364**		-0.306*
		(0.213)		(0.140)		(0.167)		(0.157)
		[-0.150]		[-0.174]		[-0.137]		[-0.120]
<u>Higher education</u> (Baseline: No)		0.160		0.212		0.155		0.362*
		(0.132)		(0.142)		(0.158)		(0.187)
		[0.062]		[0.082]		[0.058]		[0.142]
<u>Field of Studies</u> (Baseline: Fundamental Science)								
Economics		-0.231		-0.171		0.028		-0.356
		(0.248)		(0.229)		(0.273)		(0.255)
		[-0.092]		[-0.068]		[0.010]		[-0.135]
Humanities		-0.846***		-0.603**		-0.306		-0.342
		(0.275)		(0.251)		(0.294)		(0.269)
		[-0.317]		[-0.226]		[-0.116]		[-0.130]
Other		-0.606*		-0.302		-0.215		-0.398*
		(0.353)		(0.302)		(0.238)		(0.234)
		[-0.236]		[-0.119]		[-0.081]		[-0.152]
Observations	163	163	194	194	211	211	190	190
LR chi2	8.224	69.306	7.975	57.038	14.631	29.383	17.593	45.658
Prob > χ^2	0.004	0.000	0.005	0.000	0.000	0.000	0.000	0.000
Bootstrapped p-value	0.011	0.001	0.010	0.002	0.001	0.000	0.000	0.000

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses. Marginal effects estimated at the means of covariates in brackets.

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

The "Bootstrapped p-value" report the p-value associated with the group variable when applying the score cluster bootstrap correction for the small number of clusters (Kline and Santos, 2012). The bootstrap is conducted with 4999 repetitions.

Table 2.6: Politicians from the Lab v. Politicians from the Field

	HEI				HEV ^a				H ² MI				HEV ^v			
	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$
<u>Experiment</u> (Baseline: Lab)	0.033** (0.016)	0.029 (0.025)	0.202 (0.156)	0.078 (0.211)	0.043** (0.017)	0.043** (0.020)	0.067 (0.147)	0.034 (0.172)	-0.009 (0.023)	-0.006 (0.022)	0.081 (0.095)	0.054 (0.099)	0.004 (0.022)	-0.007 (0.018)	0.082 (0.129)	0.078 (0.154)
<u>Age</u>		0.004 (0.003)		0.006 (0.019)	0.004 (0.004)		-0.006 (0.018)		-0.001 (0.002)		-0.007 (0.010)		0.004* (0.002)		0.082 (0.129)	0.078 (0.154)
<u>Gender</u> (Baseline: Male)		0.070 (0.090)		-0.393 (0.339)	0.070 (0.083)		-0.426* (0.225)		-0.025 (0.054)		-0.580** (0.235)		0.023 (0.069)		0.082 (0.129)	0.078 (0.154)
<u>Higher education</u> (Baseline: No)		0.007 (0.057)		0.168 (0.304)	-0.006 (0.056)		0.402 (0.281)		-0.030 (0.040)		0.224 (0.225)		-0.029 (0.042)		0.082 (0.129)	0.078 (0.154)
<u>Field of Studies</u> (Baseline: Fundamental Science)																
<u>Economics</u>		-0.061 (0.102)		-0.157 (0.344)	-0.071 (0.077)		-0.273 (0.255)		0.095* (0.053)		0.155 (0.353)		-0.088* (0.050)		0.082 (0.129)	0.078 (0.154)
<u>Other</u>		-0.106 (0.101)		-0.758* (0.444)	-0.057 (0.093)		-0.391 (0.319)		0.055 (0.047)		-0.010 (0.293)		-0.110* (0.059)		0.082 (0.129)	0.078 (0.154)
<u>Humanities</u>									0.145* (0.077)		0.075 (0.433)		0.006 (0.096)		0.082 (0.129)	0.078 (0.154)
Observations	78	78	76	76	97	97	95	95	139	139	124	124	139	139	115	115
LR chi2	4.362	9.632	1.677	20.392	6.515	12.014	0.209	24.892	0.157	9.849	0.732	14.465	0.029	10.942	0.411	33.138
Prob > χ^2	0.037	0.141	0.195	0.002	0.011	0.062	0.648	0.000	0.692	0.197	0.392	0.044	0.864	0.141	0.522	0.000
Bootstrapped p-value	0.098	0.321	0.202	0.731	0.024	0.071	0.659	0.837	0.687	0.791	0.371	0.593	0.869	0.697	0.527	0.613

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses. Marginal effects estimated at the means of covariates in brackets.

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

The "Bootstrapped p-value" report the p-value associated with the group variable when applying the score cluster bootstrap correction for the small number of clusters (Kline and Santos, 2012). The bootstrap is conducted with 4999 repetitions.

2.4b): we fail to reject the null hypothesis of zero coefficient only when we restrict our attention to subjects with HEI ≥ 0.51 and when no control variables are introduced (see the first two columns for every restriction strategy). In all other cases the effect is statistically significant at least at the 10% level (and generally at the 5% level). In particular, being a politician decreases by about 15% to 24% the likelihood to have efficiency-oriented preferences ($\hat{\rho}_s > 0$) on average.

The third column for every restriction strategy in table 2.4 deserves some special comments. Such columns correspond to the introduction of political preferences in the models. We do not observe much of an impact as compared to second column models in the regressions on $\hat{\alpha}_s$ (table 2.4a). On the contrary, we observe a reduction in both the magnitude and the statistical significance of the "Pol" variable's coefficient in the regressions on $\hat{\rho}_s > 0$ (table 2.4b). Specifically, the effect still remains negative but the significance is dropped for subjects-based restrictions and does not resist the bootstrap correction on standard errors for observations-based

restrictions. At the same time, the impact of political preference is positive and highly significant in all specifications but the model with the HEV^v restriction. In other words, right-leaning subjects are more likely to be inclined towards efficiency.

However, the introduction of political preferences coincides with reducing the sample of Politicians by half, which complexifies the interpretation. Indeed, questions about political preferences were introduced during the end-of-experiment questionnaire, but only to Politicians from the Field and Non-Students.⁶⁷ In addition, our sample of Politicians appear slightly more left-leaning than our comparable control group, so that the two variables are weakly correlated. As a consequence, we may be confronted to an additional power issue.

In order to separate the impact of political preferences from the impact of the reduction in number of observations, we repeat the same analysis without the political preferences (tables B.12 and B.13 in appendix B.5.2). First, the coefficient on the binary group variable for each restriction strategy does not change much as compared to the ones exposed in table 2.4. Second, though the p-values indicate slightly more significant effects for HEV^a and H²MI, the gap is not sizeable with the p-values displayed here (we find at best a 10%-significant difference). Similarly, the effect on $\hat{\alpha}_s$ is virtually identical. Ultimately, that we still detect an effect while losing roughly a quarter/third observations is not anecdotal. From these observations, we conclude that a missing political preferences issue is unlikely to be an important driver of our results in the regressions with control variables and full sample of politicians.⁶⁸

Concerning the comparisons between Politicians and Students, all coefficients are

⁶⁷In particular, we avoided asking Politicians from the Lab about their political preferences because it could have negative external effects on both the recruitment (part of subjects participated out of word-to-mouth communication with past subjects) and future studies (the support of the Association des Maires de France was critical to the success of the experiment). Indeed, such questions may feel inappropriate to participants, especially politicians, and impair their perception of this kind of study. On the contrary, the recruitment process was completely independent across field sessions, therefore weakening the issue along both dimensions.

⁶⁸A more detailed analysis on the impact of political preferences on both parameters is proposed in the sections B.5.3 and B.5.5. In particular, we check whether our results are sensitive to a change in the definition of political preferences. We uncover very similar results in both sign, magnitude and significance of the binary group variable for both estimated parameters across all specifications. In addition, the results from specification curves in section B.5.1 show that we obtain more robust results on $\hat{\rho}_s > 0$ for observations-based restrictions if we change the exclusion threshold to 0.95 or if we consider Non-Linear-Least-Squares instead of Maximum Likelihood to estimate the parameters.

consistent and highly significant no matter the underlying restriction strategy (table 2.5). In the table 2.5a, the negative value of the estimated coefficient associated with the "Pol" variable in all columns indicates that Politicians are much less selfish than Students. Similarly, the negative value of the estimated coefficient in the columns of the table 2.5b indicates that Politicians are much more likely to be inclined towards equality as compared to Students.

Finally, we observe little differences between our Politicians' sub-groups (table 2.6). Concerning $\hat{\alpha}_s$, we do observe a statistically significant effect when implementing subjects-based restrictions: Politicians from the Field are more selfish than Politicians from the Lab. However it resists the introduction of control variables only when removing subjects based on the average element in the HEV.⁶⁹ Furthermore, we also observe that sign of the coefficient is not consistent across models. Concerning $\hat{\rho}_s$, though Politicians from the Field appear slightly more efficiency-oriented than Politicians from the Lab, none of the coefficients are statistically different from zero.⁷⁰

6 Discussion

The present studies focuses on testing one crucial assumption of economic models of political decision-making: that the decision-maker is self-interested. To do so, we rely on an experiment with actual local politicians in France that allows us to finely disentangle between two components of observed behavior: the Selfishness/Altruism tradeoff and the Equality/Efficiency tradeoff. Based on revealed preference techniques and the use of a structural utility model, we estimate individual parameters identifying both tradeoffs. We find that politicians display both a higher concern for others (altruism) and a higher concern for reducing inequalities

⁶⁹However, in the model with control variables, the χ^2 test on the overall fit of the model fails to reject the null hypothesis that all coefficients equal zero.

⁷⁰We further investigate the possible sources of the heterogeneity in distributional concerns within our unique sample of politicians in appendix B.5.6. Specifically, we regress estimated parameters on individual characteristics specific to politicians as well as on characteristics of the underlying city of election. We fail to uncover statistically meaningful relationships but still highlight some pointers that could be of interest for future research, in particular related to experience in political decision-making, the size of the municipality and the average income within the municipality.

(equality) than a student control group and a non-student control group designed to match politicians' observables. We also find that these results are robust to the inclusion of standard control variables, and are not driven by differences in closeness to utility maximization or by differences in prediction errors.

We acknowledge that our results may be confounded by several factors. The experimental procedures may indeed have influenced politicians' behavior. This is an issue in every experiment, but it may interact with the subject pool under scrutiny in unpredictable ways, which makes it harder to handle *ex ante*. In particular, the issue then is that we cannot identify the cause of differences in observed behavior across groups: either it corresponds to true differential in preferences or it reflects the differentiated impact of the experimental procedures on the groups.

First, since politicians are public figures, they are likely to pay specific attention to what other people think about them, which includes the experimenter. Besides, politicians may also partly think in terms of group reputation and be eager not to contribute to shaping a bad reputation for politicians in general, given that scientific results will be publicly disclosed at some point (which is common knowledge). In other words, politicians have a higher interest in signaling themselves as "moral" to the experimenter (or even to themselves).⁷¹

Considering that behaving selfishly is generally not morally praised, we could henceforth expect politicians to behave more altruistically than non-politicians because they care more about others' perception of themselves.⁷² In addition, such a moral signaling may interact with the specific payment scheme that we implement. Indeed, though we do not observe that our payment scheme have an effect on the be-

⁷¹Of course, the anonymity condition is supposed to mitigate such a signaling effect. Indeed, signaling that one is moral to another person which cannot identify yourself or with whom one can expect only little future social interaction is of limited interest. Yet, it is impossible to rule it out completely. Another design that would vary the level of public scrutiny on politicians' decisions could thus be helpful to disentangle preferences-based behavior from perception-based behavior. One should mention here the result on social image concern obtained by Banerjee et al. (2020) with local politicians in India: non-politicians behave more altruistically when decisions are public than when they are concealed, whereas no such change is observed with politicians. The authors mention an "entitlement effect" to explain such a counter-intuitive result for politicians. If it is to hold in other settings, such a finding would mean that our results are ultimately conservative.

⁷²However, such a reasoning does not *a priori* have an incidence on the Equality/Efficiency tradeoff since both equality and efficiency are equally-valuable moral principles.

havior of students (as compared to Fisman et al. (2007)'s sample), it is still possible that there is an effect for groups which are more sensitive to moral signaling. Again, we would expect politicians to behave more altruistically due to the reduction in payoffs' saliency.

However, there is no reason to assume that such an effect would be restricted to specific values along the Selfishness/Altruism tradeoff. We would indeed expect the shock to be homogeneously distributed downward: the distribution of the estimated parameter should be shifted to the left. Ultimately, this is not what we observe: Politicians are on average less selfish than Non-Students but the explanation lies on a lower proportion of near-selfish subjects mostly compensated by a bigger proportion of fair-minded subjects. Besides, the proportion of subjects weighting the recipient more than themselves is virtually identical between Politicians and Non-Students.

Second, our experiment may have prompted some in-group/out-group bias within our politician sample. Indeed, Politicians from the Lab knew that they had been recruited because they were politicians, so that they could infer that the other subjects in the rooms were politicians too. It was even more conspicuous in the field because sessions were directly organized by intermunicipal governments' and occurred in their facilities. Though it is also possible that subjects from our control groups also knew that they shared some identity with the remainder of their sessions, it is plausibly less obvious.

The effect of an in-group bias on distributional preferences is directly addressed in Müller (2019), relying on the very same methodology as ours. He does show evidence of an in-group bias. However, our results are not truly consistent with his findings. Regarding the Selfishness/Altruism tradeoff, he finds that subjects in the in-group treatment behave more altruistically, but that it is not driven by the number of nearly selfish subjects ($\hat{\alpha}_s \geq 0.9$). We report a similar difference between Politicians and both our control groups on the mean, but the difference is clearly driven by the number of near-selfish subjects. Regarding the Equality/Efficiency tradeoff, he finds an effect on the variance of the estimated parameter but not on the median, though in-group subjects appear less efficiency-oriented. On the contrary, the variance does not seem much affected in our data, since the distribution of $\hat{\rho}_s$ is relatively consistent throughout for all groups. The main difference is the distribution being above for Politicians as compared to both control groups, hence indicating a higher concern for equality over efficiency. Furthermore, had an in-group bias exists in our

data, it should have been even more salient for Politicians from the Field who share an affiliation to the same intermunicipal government. However, we do not observe much differences between the two politicians' sub-samples, and the differences lie in the opposite direction: Politicians from the Field are slightly more selfish and more efficiency-oriented than Politicians from the Lab.

Third, our results indicate that politicians' intrinsic motivations matter in their decision-making. Yet, one key element in the design is that decisions are incentivized, meaning that subjects are best placed to reveal their preferences. We acknowledge that incentives, which is a form of extrinsic motivations, may crowd out intrinsic motivations (Bénabou and Tirole, 2003). It is even the purpose of incentives in an experiment to do so (Read, 2005). Though it may not interfere with what the experiment is measuring since distributional preferences ultimately rely on monetary outcomes, it may affect politicians' motivation in participating in the experiment. Indeed, participants, the AMF and the intermunicipal governments all showed some doubts or even reluctance when they were informed about the payment. These doubts are partly responsible for our choice to implement the HRIS. Paying politicians may indeed conflict with their perception of their function in society. Indeed, the "corrupt politician" stereotype is skin-deep and the very fact that they get paid in the experiment may have been perceived as an reflection of such stereotype. In other words, they may have felt insulted.⁷³

In addition, local politicians in France are most of the time volunteers that do not receive a salary for their public services. This may affect our results in several ways. First, there may be a selection effect, with an over-representation of extrinsically-motivated politicians willing to participate. Such a bias is however not a real issue since we still observe a significant display of intrinsic motivations in our data. Second, it may introduce some cognitive dissonance between their willingness to do good in helping scientific research and the negative emotions associated with the payment. Such a cognitive dissonance would likely translate into a higher proportion of subjects violating revealed preference axioms. This is not the case as politicians do not fare worse than Non-Students in this regard. Third, it may

⁷³We note that such a repulsion for getting paid, which we came across several times during cheap talks with subjects, is also an interesting finding *per se*. Generalizing such a finding and testing alternative payment schemes could be an important follow-up.

also lead subjects to avoid the strictly selfish and over-report the strictly altruistic choice. This would represent an effort of defiance with the feeling that the payment made them have. We do observe politicians to be less strictly selfish. But such an explanation do not account for the focus on fair-mindedness. Besides, we do not observe that politicians are more altruistic than Non-Students.

We consequently believe that one should be very careful to these various elements when running experiments with politicians. Nevertheless, we also believe that the impact on the results is fairly limited, given the patterns that we observe.

We also briefly comment on our results regarding the Equality/Efficiency trade-off. As previously mentionned, what we find contradicts our working hypotheses which pointed towards a prevalence of efficiency over equality within a pool of politicians, or at least no difference with respect to a comparable control group. However, our working hypotheses were formulated based on existing studies whose characteristics differ from ours in two important ways.

First, the majority of studies happened in developping countries whereas ours was run in France. In developping countries, achieving equality within the society, which is not an essential good, may not be a primary objective of policy-makers as opposed to increasing the nation's wealth. On the contrary, countries who already have reached an advanced state of economic development may be more concerned about reducing inequalities, and so would be their political leaders.

Second, our study focuses on local-level politicians, which is not the case of all previous studies which may include upper-tiers politicians. Focusing on local-level politicians in France is interesting because they are socio-demographically more comparable to the remainder of the French population than upper-tiers politicians. Furthermore, many French upper-tiers politicians start their career at the municipality level. However, politicians' distributional preferences may differ depending on their administrative layer: deciding in the name of a big city, a region or even a whole nation instead of a small town may modify how politicians trade-off equality and efficiency. In particular, politicians facing a bigger territory may be more inclined towards maximizing the overall payoffs than reducing inequalities.

7 Concluding remarks

Our results have several implications.

First, our results entail that intrinsic motivations are key to the understanding of politicians' behavior at the individual level. As emphasized by Hume's introductory quote, one influential approach to the design of political institutions relies on the necessity to assume self-interested politicians in order to build constitutional rules and incentive schemes that guards against public good's hijacking and renders behavior predictable (Kliemt, 2005). However, extrinsic rewards can be counterproductive with intrinsically-motivated agents (Bénabou and Tirole, 2003, 2006) or may lead to the selection of extrinsically-motivated individuals only (Besley, 2005). The importance of intrinsic motivations for politicians may consequently explain the mixed findings regarding the impact of pay rises on the quality of the politicians' pool (Braendle, 2016, Dal Bó and Finan, 2018).⁷⁴ At the same time, that politicians are motivated by other-regarding preferences appears valuable in representative democracies as it may drive representatives to cater to citizens' preferences even though they are not externally incentivized to do so.

Second, and relatedly, theoretical models of political decisions which traditionally rely on a self-interest assumption are likely to benefit from incorporating distributional preferences in terms of predictive accuracy.

Third, our results suggest that politicians's distributional preferences are not aligned with the distributional preferences of the underlying population.⁷⁵ In a representative democracy with an imperfect control of the represented on their representatives, such a finding ultimately entails that political decisions are likely not to match the will of the people, even when politicians only behave out of what

⁷⁴In a framework where politicians vary in both competence and public service motivation, Fedele and Giannoccolo (2020) find that paying politicians is effective only for intermediate wage values due to the crowding out of either skilled or highly-motivated candidates if the pay is too low or too high.

⁷⁵We acknowledge that our experiment do not directly compare politicians' distributional preferences with the distributional preferences of the very people that elect them. Such an endeavour would have ideally required surveying citizens in every town from which one politician agreed upon participating. Alternatively, national surveys could have been useful, though noisier. However, we can already say something about what the comparison with citizens may look like. Indeed, our control group is likely to be more altruistic and more prone to favor equality over efficiency than the remainder of the (younger) population. As a consequence, the gap between politicians and the population in terms of both tradeoffs is likely to be even greater than the one we observe.

they consider as "good". Indeed, when citizens would favor increasing total income, even at the expense of a more equal society, politicians lean more towards reducing inequalities.

Fourth, our experiment shed some light on the usefulness of recruiting students in experiments designed to tackle political issues. Indeed, actual politicians appear even farther away from students than non-students do in terms of distributional concerns. Such a finding prompts some caution when extrapolating from experimental results on political decisions obtained with convenience student samples.

Fifth, our results question the way economists are used to "whisper in the ears of the prince" (Roth, 1986). Indeed, economics is essentially driven by the search for an efficient allocation of resources, so that scholars' public policy recommendations are generally based on emphasizing how to reach such efficient allocation. Then, economists are often annoyed by the difficulty that their fancy welfare-maximizing solutions face in actually being adopted by policy-makers. Our results suggest a possible explanation: politicians are simply not responsive to efficiency talks. Scholars who wish to convince the prince to adopt their recommendations may thus consider alternative formulations to enhance the transmission of the message based on what the prince is most sensitive to.

We conclude by emphasizing some limitations of the present study and outline avenues for future research. We start with some external validity comments.

Our design holds the highly-valuable feature that it enables to finely measure individual-level distributional preferences in a very simple environment. However, key elements of the actual political decision-making environment are missing which may introduce some discrepancies between behavior elicited in the lab and real-world behavior. Actual political decision-making indeed shows several collective dimensions. In particular, politicians have been elected to fulfil voters' preferences. It could thus be interesting to explore the interplay between politicians distributional concerns and voters' preferences, especially when they are in conflict.

Relatedly, the definition of self-interest that we adopt may not faithfully represent the actual incentives of politicians. In practice, politicians' self-interest is often not directly materialistic. In particular, the objective of politicians is generally to get elected (which may ultimately translate into some materialistic benefits). Furthermore, actual political decision-making is a collective product: politicians ultimately vote for any given decision they make, which introduces some strategic

elements into the decision-making process which may not be neutral with respect to distributional concerns. Running a distribution experiment that would introduce some variations in these collective dimensions with a sample of politicians could consequently provide interesting insights as to the way actual policy-makers behave.

More generally, that we observe a specific profile in terms of distributional preferences for politicians does not mean that they act upon their preferences when making a political decision. It would consequently be an important inquiry to correlate the distributional preferences of policy-makers with implemented policies.

Lastly, replicating our results on a larger scale would also prove highly valuable. It would indeed allow cross-regional comparisons to identify how different political, social, economic demographic or even geographic backgrounds interplay with politicians' preferences, as a generalization of the small-scale heterogeneity analysis proposed in the appendices of this chapter.

Additionally, the present study does not allow to identify the source of the difference between politicians and a comparable group. Aside from an omitted variable issue, such a difference can be driven either by a pre-office selection effect or by an institutional effect once in office. In order to disentangle both effects, one would need to compare politicians with a valid counterfactual group, namely political candidates who failed at getting elected, in addition to comparing politicians with the remainder of the population.⁷⁶ For instance, one could exploit the discontinuity introduced by close races between successful (elected) and unsuccessful (unelected) candidates, as in Enemark et al. (2016). Yet, recovering the set of political candidates is often even more challenging than collecting data from successful politicians. Indeed, the set of "unsuccessful candidates" not only includes running candidates that were not elected but also the set of "unseen candidates" (Fowler and McClure, 1989) that did not enter the race. Alternatively, one could run large-scale surveys to identify political candidates or politically-ambitious citizens within the general population. But such an endeavour is likely costly because the share of political candidates in the total population is extremely small, causing a "rare-event" problem (Gulzar, 2021).

When inspecting the mechanisms behind our results, we must acknowledge that

⁷⁶Indeed, our political experience variable is probably too crude to isolate a selection-into-politics effect from a political-institutions effect on preferences.

our findings on the Equality/Efficiency tradeoff contradict our literature-based conjectures. Although we hypothesized that politicians would be more prone to favoring Efficiency, we observe the opposite. Together with the fact that politicians are more fair-minded than our control groups, such a result points towards a very strong focal point in preferences. In other words, politicians seem obsessed with income equalization. Of course, the discrepancy with existing evidence could stem from the fact that our study's design sufficiently differ from the literature, which is only burgeoning. However, if we were to consider our result as robust, which we believe it is given the additional tests that we provide, we may suggest two possible explanations. First, politicians are elected by voters under circumstances when they are uncertain about their future income positions. In such a case, voters may be willing to elect pro-social individuals that would care for the well-being of every "position" in society, in case they end up having a bad one. This joins the theory of constitutional government where citizens placed under the veil of ignorance are driven out of their self-interest to consider all situations in society (Buchanan and Tullock, 1962). Second, since political decisions are achieved through a collective process, they are confronted with both a collective action problem and a coordination issue. As emphasized by Jankowski (2019), the selection of intrinsically-motivated (or pro-social) individuals may help solve the former. As for the latter, it certainly helps to have decision-makers endowed with a focal point in preferences that makes them spontaneously agree that everybody has to earn something from the decision.

References

- Afriat, S. N. (1967). The construction of utility functions from expenditure data. International Economic Review, 8(1):67–77.
- Afriat, S. N. (1972). Efficiency estimation of production functions. International Economic Review, pages 568–598.
- Afriat, S. N. (1973). On a system of inequalities in demand analysis: an extension of the classical method. International Economic Review, pages 460–472.
- Alekseev, A., Alm, J., Sadiraj, V., and Sjoquist, D. L. (2021). Experiments on the fly. Journal of Economic Behavior & Organization, 186:288–305.
- Alesina, A. and Angeletos, G.-M. (2005). Fairness and redistribution. American Economic Review, 95(4):960–980.
- Alesina, A. and Giuliano, P. (2011). Preferences for redistribution. In Handbook of social economics, volume 1, pages 93–131. Elsevier.
- Alesina, A., Glaeser, E., and Glaeser, E. L. (2004). Fighting poverty in the US and Europe: A world of difference. Oxford University Press.
- Almås, I., Cappelen, A. W., and Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? Journal of Political Economy, 128(5):1753–1788.
- Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. Econometrica, 70(2):737–753.
- Bali, V. A., Robison, L. J., and Winder, R. (2020). What motivates people to vote? the role of selfishness, duty, and social motives when voting. SAGE Open, 10(4):2158244020950376.
- Baltussen, G., Post, G. T., Van Den Assem, M. J., and Wakker, P. P. (2012). Random incentive systems in a dynamic choice experiment. Experimental Economics, 15(3):418–443.
- Banerjee, P., Sen, K., Nicolo, A., Iversen, V., and Mitra, S. (2020). What motivates politicians? evidence from a lab-in-the-field experiment in india. Working paper.

- Beatty, T. K. and Crawford, I. A. (2011). How demanding is the revealed preference approach to demand? American Economic Review, 101(6):2782–95.
- Behrman, J. R. and Birdsall, N. (1988). The equity-productivity tradeoff: Public school resources in brazil. European Economic Review, 32(8):1585–1601.
- Behrman, J. R. and Craig, S. G. (1987). The distribution of public services: An exploration of local governmental preferences. The American Economic Review, pages 37–49.
- Bel, G. and Fageda, X. (2009). Preventing competition because of ‘solidarity’: Rhetoric and reality of airport investments in spain. Applied Economics, 41(22):2853–2865.
- Bellemare, C. and Kröger, S. (2007). On representative social capital. European Economic Review, 51(1):183–202.
- Bellemare, C., Kröger, S., and Van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. Econometrica, 76(4):815–839.
- Bellemare, C., Kröger, S., and van Soest, A. (2011). Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool. Journal of Economic Behavior & Organization, 78(3):349–365.
- Ben-Ner, A., Kramer, A., and Levy, O. (2008). Economic and hypothetical dictator game experiments: Incentive effects at the individual level. The Journal of Socio-Economics, 37(5):1775–1784.
- Bénabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. The Review of Economic Studies, 70(3):489–520.
- Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. American Economic Review, 96(5):1652–1678.
- Beniers, K. J. and Dur, R. (2007). Politicians’ motivation, political culture, and electoral competition. International Tax and Public Finance, 14(1):29–54.
- Benz, M. and Meier, S. (2008). Do people behave in experiments as in the field?—evidence from donations. Experimental Economics, 11(3):268–281.

- Bernheim, B. D. and Kartik, N. (2014). Candidates, character, and corruption. American Economic Journal: Microeconomics, 6(2):205–46.
- Besley, T. (2005). Political selection. Journal of Economic Perspectives, 19(3):43–60.
- Besley, T. and Ghatak, M. (2005). Competition and incentives with motivated agents. American Economic Review, 95(3):616–636.
- Besley, T. and Ghatak, M. (2018). Prosocial motivation and incentives. Annual Review of Economics, 10:411–438.
- Best, H. (2011). Does personality matter in politics? personality factors as determinants of parliamentary recruitment and policy preferences. Comparative Sociology, 10(6):928–948.
- Blanes i Vidal, J., Draca, M., and Fons-Rosen, C. (2012). Revolving door lobbyists. American Economic Review, 102(7):3731–48.
- Boadway, R. and Keen, M. (2000). Redistribution. Handbook of income distribution, 1:677–789.
- Braendle, T. (2016). Do institutions affect citizens’ selection into politics? Journal of Economic Surveys, 30(2):205–227.
- Broockman, D. E. (2013). Black politicians are more intrinsically motivated to advance blacks’ interests: A field experiment manipulating political incentives. American Journal of Political Science, 57(3):521–536.
- Buchanan, J. M. and Tullock, G. (1962). The Calculus of Consent, volume 12. Ann Arbor: University of Michigan Press.
- Burns, D. J., Reid, J. S., Toncar, M., Fawcett, J., and Anderson, C. (2006). Motivations to volunteer: The role of altruism. International Review on Public and Nonprofit Marketing, 3(2):79–91.
- Butler, D. M. and Kousser, T. (2015). How do public goods providers play public goods games? Legislative Studies Quarterly, 40(2):211–240.
- Calvert, R. L. (1985). Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence. American Journal of Political Science, pages 69–95.

- Camerer, C. F. (2011). Behavioral game theory: Experiments in strategic interaction. Princeton University Press.
- Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics, 90(3):414–427.
- Cameron, A. C. and Miller, D. L. (2015). A practitioner’s guide to cluster-robust inference. Journal of Human Resources, 50(2):317–372.
- Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. American Economic Review, 97(3):818–827.
- Cappelen, A. W., Moene, K. O., Sørensen, E. Ø., and Tungodden, B. (2013). Needs versus entitlements—an international fairness experiment. Journal of the European Economic Association, 11(3):574–598.
- Cappelen, A. W., Nygaard, K., Sørensen, E. Ø., and Tungodden, B. (2015). Social preferences in the lab: A comparison of students and a representative population. The Scandinavian Journal of Economics, 117(4):1306–1326.
- Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Responsibility for what? fairness and individual responsibility. European Economic Review, 54(3):429–441.
- Caprara, G., Francescato, D., Mebane, M., Sorace, R., and Vecchione, M. (2010). Personality foundations of ideological divide: A comparison of women members of parliament and women voters in Italy. Political Psychology, 31(5):739–762.
- Carpenter, J., Connolly, C., and Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. Experimental Economics, 11(3):282–298.
- Carpenter, J. and Myers, C. K. (2010). Why volunteer? evidence on the role of altruism, image, and incentives. Journal of Public Economics, 94(11-12):911–920.
- Castells, A. and Solé-Ollé, A. (2005). The regional allocation of infrastructure investment: The role of equity, efficiency and political factors. European Economic Review, 49(5):1165–1205.

- Cetre, S., Lobeck, M., Senik, C., and Verdier, T. (2019). Preferences over income distribution: Evidence from a choice experiment. Journal of Economic Psychology, 74:102202.
- Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. Journal of Economic Behavior & Organization, 131:141–150.
- Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. The Quarterly Journal of Economics, 117(3):817–869.
- Chaudhuri, A., Iversen, V., Jensenius, F. R., and Maitra, P. (2020). Selecting the best of us? politician quality in village councils in west bengal, india. Technical report, CESifo Working Paper.
- Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007). Consistency and heterogeneity of individual behavior under uncertainty. American Economic Review, 97(5):1921–1938.
- Choi, S., Kariv, S., Müller, W., and Silverman, D. (2014). Who is (more) rational? American Economic Review, 104(6):1518–50.
- Clot, S., Grolleau, G., and Ibanez, L. (2018). Shall we pay all? an experimental test of random incentivized systems. Journal of Behavioral and Experimental Economics, 73:93–98.
- Cooper, D. J. and Kagel, J. H. (2016). Other-regarding preferences. The handbook of experimental economics, 2:217.
- Coughlin, P. J. (1992). Probabilistic voting theory. Cambridge University Press.
- Cox, J. C. (2004). How to identify trust and reciprocity. Games and Economic Behavior, 46(2):260–281.
- Craig, S. G. (1987). The deterrent impact of police: An examination of a locally provided public service. Journal of Urban Economics, 21(3):298–311.
- Dal Bó, E., Dal Bó, P., and Snyder, J. (2009). Political dynasties. The Review of Economic Studies, 76(1):115–142.
- Dal Bó, E. and Finan, F. (2018). Progress and perspectives in the study of political selection. Annual Review of Economics, 10:541–575.

- Dal Bó, E., Finan, F., and Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. The Quarterly Journal of Economics, 128(3):1169–1218.
- Dawes, C. T., Loewen, P. J., and Fowler, J. H. (2011). Social preferences and political participation. The Journal of Politics, 73(3):845–856.
- de Figueiredo Jr, R. J. and Edwards, G. (2007). Does private money buy public policy? campaign contributions and regulatory outcomes in telecommunications. Journal of Economics & Management Strategy, 16(3):547–576.
- Dean, M. and Martin, D. (2016). Measuring rationality with the minimum cost of revealed preference violations. Review of Economics and Statistics, 98(3):524–534.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. Journal of Personality and Social Psychology, 18(1):105.
- Delfgaauw, J. and Dur, R. (2008). Incentives and workers' motivation in the public sector. The Economic Journal, 118(525):171–191.
- Dixit, A. and Londregan, J. (1998). Ideology, tactics, and efficiency in redistributive politics. The Quarterly Journal of Economics, 113(2):497–529.
- Downs, A. (1957). An economic theory of democracy. New York: Harper & Row.
- Durante, R., Putterman, L., and Van der Weele, J. (2014). Preferences for redistribution and perception of fairness: An experimental study. Journal of the European Economic Association, 12(4):1059–1086.
- Eckel, C. C. and Grossman, P. J. (2000). Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. Experimental Economics, 3(2):107–120.
- Edlin, A., Gelman, A., and Kaplan, N. (2007). Voting as a rational choice: Why and how people vote to improve the well-being of others. Rationality and society, 19(3):293–314.
- Eggers, A. C. and Hainmueller, J. (2009). Mps for sale? returns to office in postwar british politics. American Political Science Review, 103(4):513–533.

- Enemark, D., Gibson, C. C., McCubbins, M. D., and Seim, B. (2016). Effect of holding office on the behavior of politicians. Proceedings of the National Academy of Sciences, 113(48):13690–13695.
- Engel, C. (2011). Dictator games: A meta study. Experimental Economics, 14(4):583–610.
- Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. American Economic Review, 94(4):857–869.
- Epper, T., Fehr, E., and Senn, J. (2020). Other-regarding preferences and redistributive politics. Working Paper 339, University of Zurich, Department of Economics.
- Evren, Ö. (2012). Altruism and voting: A large-turnout result that does not rely on civic duty or cooperative behavior. Journal of Economic Theory, 147(6):2124–2157.
- Fafchamps, M. and Labonne, J. (2017). Do politicians’ relatives get better jobs? evidence from municipal elections. The Journal of Law, Economics, and Organization, 33(2):268–300.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. The Quarterly Journal of Economics, 133(4):1645–1692.
- Fatas, E., Neugebauer, T., and Tamborero, P. (2007). How politicians make decisions: A political choice experiment. Journal of Economics, 92(2):167–196.
- Fearon, J. D. (1999). Electoral accountability and the control of politicians: Selecting good types versus sanctioning poor performance. In Przeworski, A., Stokes, S. C. S., Stokes, S. C., and Manin, B., editors, Democracy, accountability, and representation, volume 2, pages 55–97. Cambridge University Press.
- Fedele, A. and Giannoccolo, P. (2020). Paying politicians: Not too little, not too much. Economica, 87(346):470–489.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. The Quarterly Journal of Economics, 114(3):817–868.

- Ferraz, C. and Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Fisman, R., Jakiela, P., and Kariv, S. (2015a). How did distributional preferences change during the great recession? *Journal of Public Economics*, 128:84–95.
- Fisman, R., Jakiela, P., and Kariv, S. (2017). Distributional preferences and political behavior. *Journal of Public Economics*, 155:1–10.
- Fisman, R., Jakiela, P., Kariv, S., and Markovits, D. (2015b). The distributional preferences of an elite. *Science*, 349(6254):aab0096.
- Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, 97(5):1858–1876.
- Fisman, R., Schulz, F., and Vig, V. (2014). The private returns to public office. *Journal of Political Economy*, 122(4):806–862.
- Fligner, M. A. and Policello, G. E. (1981). Robust rank procedures for the behrens-fisher problem. *Journal of the American Statistical Association*, 76(373):162–168.
- Folke, O., Persson, T., and Rickne, J. (2016). The primary effect: Preference votes and political promotions. *American Political Science Review*, 110(3):559–578.
- Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3):347–369.
- Foucault, M. and Bono, P.-H. (2020). Enquête 2020 sur les maires de france. Technical report, CEVIPOF.
- Fowler, J. H. (2006). Altruism and turnout. *The Journal of Politics*, 68(3):674–683.
- Fowler, J. H. and Kam, C. D. (2007). Beyond the self: Social identity, altruism, and political participation. *The Journal of Politics*, 69(3):813–827.
- Fowler, L. L. and McClure, R. D. (1989). *Political ambition: Who decides to run for Congress*. Yale University Press.

- Francois, P. (2000). 'public service motivation' as an argument for government provision. Journal of Public Economics, 78(3):275–299.
- Frey, B. S. (1992). Tertium datur: Pricing, regulating and intrinsic motivation. Kyklos, 45(2):161–184.
- Frigau, L., Medda, T., and Pelligra, V. (2019). From the field to the lab. an experiment on the representativeness of standard laboratory subjects. Journal of Behavioral and Experimental Economics, 78:160–169.
- Gächter, S., Nosenzo, D., Renner, E., and Sefton, M. (2012). Who makes a good leader? cooperativeness, optimism, and leading-by-example. Economic Inquiry, 50(4):953–967.
- Gächter, S. and Renner, E. (2018). Leaders as role models and 'belief managers' in social dilemmas. Journal of Economic Behavior & Organization, 154:321–334.
- Gagliarducci, S. and Manacorda, M. (2020). Politics in the family: Nepotism and the hiring decisions of Italian firms. American Economic Journal: Applied Economics, 12(2):67–95.
- Georgellis, Y., Iossa, E., and Tabvuma, V. (2011). Crowding out intrinsic motivation in the public sector. Journal of Public Administration Research and Theory, 21(3):473–493.
- Gershberg, A. I. and Schuermann, T. (2001). The efficiency–equity trade-off of schooling outcomes: public education expenditures and welfare in Mexico. Economics of Education Review, 20(1):27–40.
- Gilens, M. and Page, B. I. (2014). Testing theories of American politics: Elites, interest groups, and average citizens. Perspectives on Politics, 12(3):564–581.
- Gniza, J., Abraham, M., Lehmann, I., and Lorek, K. (2022). Expecting equality or efficiency? a factorial survey on the justice perception of regional redistributive policies. Regional Studies, pages 1–15.
- Golden, M. and Min, B. (2013). Distributive politics around the world. Annual Review of Political Science, 16:73–99.

- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. Journal of the Economic Science Association, 1(1):114–125.
- Gulzar, S. (2021). Who enters politics and why? Annual Review of Political Science, 24:253–275.
- Gulzar, S. and Khan, M. Y. (2021). 'good politicians': Experimental evidence on motivations for political candidacy and government performance. Working paper available at [ssrn 3826067](https://ssrn.com/abstract=3826067).
- Güth, W., Schmidt, C., and Sutter, M. (2007). Bargaining outside the lab—a newspaper experiment of a three-person ultimatum game. The Economic Journal, 117(518):449–469.
- Hafner-Burton, E. M., Hughes, D. A., and Victor, D. G. (2013). The cognitive revolution and the political psychology of elite decision making. Perspectives on Politics, pages 368–386.
- Hanania, R. (2017). The personalities of politicians: A big five survey of american legislators. Personality and Individual Differences, 108:164–167.
- Harsanyi, J. C. (1975). Can the maximin principle serve as a basis for morality? a critique of john rawls's theory. American Political Science Review, 69(2):594–606.
- Heß, M., Von Scheve, C., Schupp, J., Wagner, A., and Wagner, G. G. (2018). Are political representatives more risk-loving than the electorate? evidence from german federal and state parliaments. Palgrave Communications, 4(1):1–7.
- Heufer, J. (2013). Testing revealed preferences for homotheticity with two-good experiments. Experimental Economics, 16(1):114–124.
- Heufer, J. and Hjertstrand, P. (2015). Consistent subsets: Computationally feasible methods to compute the houtman–maks-index. Economics Letters, 128:87–89.
- Heufer, J. and Hjertstrand, P. (2019). Homothetic efficiency: Theory and applications. Journal of Business & Economic Statistics, 37(2):235–247.
- Hines, J. R. and Thaler, R. H. (1995). The flypaper effect. Journal of Economic Perspectives, 9(4):217–226.

- Jacobs, L. R. and Page, B. I. (2005). Who influences us foreign policy? American Political Science Review, 99(1):107–123.
- Jakiela, P. (2013). Equity vs. efficiency vs. self-interest: on the use of dictator games to measure distributional preferences. Experimental Economics, 16(2):208–221.
- Jakiela, P. (2015). How fair shares compare: Experimental evidence from two cultures. Journal of Economic Behavior & Organization, 118:40–54.
- Janezic, K. A. and Gallego, A. (2020). Eliciting preferences for truth-telling in a survey of politicians. Proceedings of the National Academy of Sciences, 117(36):22002–22008.
- Jankowski, R. (2007). Altruism and the decision to vote: Explaining and testing high voter turnout. Rationality and Society, 19(1):5–34.
- Jankowski, R. (2019). Altruism and political participation. In Congleton, R. D., Grofman, B., and Voigt, S., editors, The Oxford Handbook of Public Choice, volume 1, pages 352–361. Oxford Handbooks.
- Joly, J., Soroka, S., and Loewen, P. (2019). Nice guys finish last: Personality and political success. Acta Politica, 54(4):667–683.
- Kam, C. D., J, C. S., and Loewen, J. H. (2019). Altruism, participation, and political context. In Loewen, J. H. and Rubenson, D., editors, Duty and choice: the evolution of the study of voting and voters. University of Toronto Press.
- Kemmerling, A. and Stephan, A. (2010). The determinants of regional transport investment across europe. In The Political Economy of Inter-Regional Fiscal Flows. Edward Elgar Publishing.
- Kerschbamer, R. (2015). The geometry of distributional preferences and a non-parametric identification approach: The equality equivalence test. European Economic Review, 76:85–103.
- Kerschbamer, R. and Müller, D. (2020). Social preferences and political attitudes: An online experiment on a large heterogeneous sample. Journal of Public Economics, 182:104076.

- Kertzer, J. D. and Renshon, J. (2022). Experiments and surveys on political elites. Annual Review of Political Science, 25:1–26.
- Kirchgässner, G. (2014). The role of homo oeconomicus in the political economy of james buchanan. Constitutional Political Economy, 25(1):2–17.
- Kliemt, H. (2005). Public choice and political philosophy: Reflections on the works of gordon spinoza and david immanuel buchanan. Public Choice, 125(1):203–213.
- Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. Journal of Econometric Methods, 1(1):23–41.
- Knack, S. (1992). Social altruism and voter turnout: Evidence from the 1991 nes pilot study. 1991 NES Pilot Study Reports.
- Konow, J. (2003). Which is the fairest one of all? a positive analysis of justice theories. Journal of Economic Literature, 41(4):1188–1239.
- Kuliomina, J. (2021). Do personal characteristics of councilors affect municipal budget allocation? European Journal of Political Economy, 70:102034.
- LeVeck, B. L., Hughes, D. A., Fowler, J. H., Hafner-Burton, E., and Victor, D. G. (2014). The role of self-interest in elite bargaining. Proceedings of the National Academy of Sciences, 111(52):18536–18541.
- Lewis, G. B. and Frank, S. A. (2002). Who wants to work for the government? Public Administration Review, 62(4):395–404.
- Lindbeck, A. and Weibull, J. W. (1987). Balanced-budget redistribution as the outcome of political competition. Public Choice, 52(3):273–297.
- Linde, J. and Vis, B. (2017). Do politicians take risks like the rest of us? an experimental test of prospect theory under mps. Political Psychology, 38(1):101–117.
- Lomasky, L. (2008). Swing and a myth: A review of caplan’s” the myth of the rational voter”. Public Choice, 135(3/4):469–484.
- Mansbridge, J. (2003). Rethinking representation. American Political Science Review, pages 515–528.

- Mas-Colell, A., Whinston, M. D., Green, J. R., et al. (1995). Microeconomic theory, volume 1. Oxford university press New York.
- McCrain, J. (2018). Revolving door lobbyists and the value of congressional staff connections. The Journal of Politics, 80(4):1369–1383.
- McMillan, J. and Zoido, P. (2004). How to subvert democracy: Montesinos in peru. Journal of Economic Perspectives, 18(4):69–92.
- Müller, D. (2019). The anatomy of distributional preferences with group identity. Journal of Economic Behavior & Organization, 166:785–807.
- Okun, A. M. (1975). Equality and Efficiency: The Big Tradeoff. Washington, DC: The Brookings Institution.
- Pedersen, L. H. (2014). Committed to the public interest? motivation and behavioural outcomes among local councillors. Public Administration, 92(4):886–901.
- Perry, J. L. and Wise, L. R. (1990). The motivational bases of public service. Public Administration Review, pages 367–373.
- Prendergast, C. (2007). The motivation and bias of bureaucrats. American Economic Review, 97(1):180–196.
- Rainey, H. G. and Steinbauer, P. (1999). Galloping elephants: Developing elements of a theory of effective government organizations. Journal of Public Administration Research and Theory, 9(1):1–32.
- Rawls, J. (1971). A theory of justice. Cambridge, Harvard University Press.
- Read, D. (2005). Monetary incentives, what are they good for? Journal of Economic Methodology, 12(2):265–276.
- Reinikka, R. and Svensson, J. (2004). Local capture: evidence from a central government transfer program in uganda. The Quarterly Journal of Economics, 119(2):679–705.
- Ritz, A. (2015). Public service motivation and politics: Behavioural consequences among local councillors in switzerland. Public Administration, 93(4):1121–1137.

- Robalo, P., Schram, A., and Sonnemans, J. (2017). Other-regarding preferences, in-group bias and political participation: An experiment. Journal of Economic Psychology, 62:130–154.
- Roth, A. E. (1986). Laboratory experimentation in economics. Economics & Philosophy, 2(2):245–273.
- Schnellenbach, J. and Schubert, C. (2015). Behavioral political economy: A survey. European Journal of Political Economy, 40:395–417.
- Schurr, A. and Ritov, I. (2016). Winning a competition predicts dishonest behavior. Proceedings of the National Academy of Sciences, 113(7):1754–1759.
- Scott, C. and Medeiros, M. (2020). Personality and political careers: What personality types are likely to run for office and get elected? Personality and Individual Differences, 152:109600.
- Selten, R. (1991). Properties of a measure of predictive success. Mathematical Social Sciences, 21(2):153–167.
- Sheffer, L., Loewen, P. J., Soroka, S., Walgrave, S., and Sheafer, T. (2018). Non-representative representatives: An experimental study of the decision making of elected politicians. The American Political Science Review, 112(2):302–321.
- Shoup, C. S. (1989). Rules for distributing a free government service among areas of a city. National Tax Journal, 42(2):103–121.
- Simon, H. A. (1995). Rationality in political behavior. Political Psychology, pages 45–61.
- Solé-Ollé, A. (2013). Inter-regional redistribution through infrastructure investment: tactical or programmatic? Public Choice, 156(1):229–252.
- Stahl, D. O. and Haruvy, E. (2006). Other-regarding preferences: Egalitarian warm glow, empathy, and group size. Journal of Economic Behavior & Organization, 61(1):20–41.
- Stantcheva, S. (2021). Understanding tax policy: How do people reason? The Quarterly Journal of Economics, 136(4):2309–2369.

- Traub, S., Seidl, C., and Schmidt, U. (2009). An experimental study on individual choice, social welfare, and social preferences. European Economic Review, 53(4):385–400.
- Truex, R. (2014). The returns to office in a “rubber stamp” parliament. American Political Science Review, 108(2):235–251.
- Varian, H. R. (1982). The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society, pages 945–973.
- Varian, H. R. (1983). Non-parametric tests of consumer behaviour. The Review of Economic Studies, 50(1):99–110.
- Vis, B. and Stolwijk, S. (2020). Conducting quantitative studies with the participation of political elites: best practices for designing the study and soliciting the participation of political elites. Quality & Quantity, pages 1–37.
- Vollan, B., Blanco, E., Steimanis, I., Petutschnig, F., and Prediger, S. (2020). Procedural fairness and nepotism among local traditional and democratic leaders in rural namibia. Science Advances, 6(15):eaay7651.
- Wallerstein, M. (2004). Behavioral economics and political economy. Nordic Journal of Political Economy, 30:37–48.
- Weber, M. (1918). Politik als beruf, gesammelte politische schriften (muenchen, 1921). Originally a speech at munich university.
- Wilson, R. K. (2011). The contribution of behavioral economics to political science. Annual Review of Political Science, 14:201–223.
- Wittman, D. (1977). Candidates with policy preferences: A dynamic model. Journal of Economic Theory, 14(1):180–189.
- Zamir, E. and Sulitzeanu-Kenan, R. (2018). Explaining self-interested behavior of public-spirited policy makers. Public Administration Review, 78(4):579–592.

Appendix B

The Distributional Preferences of Politicians: Experimental Evidence

Contents

B.1 Complementary revealed preference analysis	214
B.2 Complements on the empirical strategy	226
B.3 Complementary analyses on group comparisons from raw data .	235
B.4 Complementary analyses on group comparisons from parameters	243
B.5 Complementary regression analyses	263
References	287

B.1 Complementary revealed preference analysis

In this section, we detail the revealed preference tests that we perform on our data. These tests provide a scientific approach for designing the restrictions on subjects or observations that we use in the subsequent structural estimation process.

One major requirement for estimating an utility function is that observed behavior is indeed compatible with an utility maximization framework. Classical revealed preference theory provides a straightforward test based on the seminal theorem by Afriat (1967) (and further refined by Varian, 1982): given a finite collection of budget sets, observed decisions can be rationalized by a well-behaved (piecewise linear, continuous, monotonically increasing and concave) utility function if and only if they comply with the Generalized Axiom of Revealed Preference (GARP). Building on the large number of decisions and price variations supplied by our experiment, we implement the GARP test and compute several rationality indices indicating how far each subject deviates from the axiom in the chapter 1. From such analysis, we concluded that the very large majority of subjects fall close to utility maximization.

The usual next step is to argue that such a result warrants estimating a CES utility function to recover individual parameters identifying distributional preferences. However, Afriat (1967)'s theorem does only tell us that there exists some well-behaved utility function for datasets abiding by GARP. It does not provide more detailed information about such utility function's functional form. Though particularly convenient for our purpose (see below), the CES utility function actually comes with additional hypotheses concerning the underlying preferences. In particular, it implies that preferences are homothetic. A function is said to be "homothetic" if it is a monotonic transformation of another function that is homogeneous of degree 1. In other words, a homothetic utility maximizer that is willing to buy π_s^i at prices p_s^i will also prefer $\lambda\pi_s^i$ when prices are constant but income is multiplied by λ . In practice, such an hypothesis is rarely tested (or even mentioned explicitly).

Fortunately, the literature identifies some ways to test for inconsistencies with homothetic preferences analogously to the tests for inconsistencies with well-defined preferences. Specifically, theorem by Varian (1983) states that observed decisions can be rationalized by a (well-behaved) homothetic utility function if and only if they

comply with the Homothetic Axiom of Revealed Preference (HARP). HARP is satisfied if, for all distinct choices of indices i, j, \dots, l we observe $(\mathbf{p}_s^i \pi_s^j)(\mathbf{p}_s^j \pi_s^k) \dots (\mathbf{p}_s^l \pi_s^i) \geq B_s^i B_s^j \dots B_s^l = 1$.

In a two-commodities space, such an axiom boils down to $(\mathbf{p}_s^i \pi_s^j)(\mathbf{p}_s^j \pi_s^i) \geq 1 \forall i, j$ (Heufer, 2013), which is straightforward to implement.

However, the HARP test is an exact test which subjects either pass or fail. Since subjects are likely to make small mistakes (e.g., due to computation mistakes or fatigue), we rather focus on several continuous measures of closeness to homothetic utility maximization. The main one is the Homothetic Efficiency Index (HEI, see Heufer, 2013, Heufer and Hjertstrand, 2019), which quantifies the amount by which all budget lines must be shifted in order to remove all HARP violations. The HEI is quite efficient in summarizing into a single number the information about the extent of HARP violations for a given subject. However, it is ultimately a lower bound on homothetic efficiency and it does not provide information as to which observations are causing the violations. It may indeed appear rather punitive that a subject be considered a massive violator when he deviated only once (but substantially), as opposed to a repeat violator (but by small amounts).

Consequently, we also compute two additional measures to consider alternative formulations of "closeness" to homothetic utility maximization: the Homothetic Houtman-Maks Index (H²MI, see Heufer and Hjertstrand, 2015) and the Homothetic Efficiency Vector (HEV, see Heufer and Hjertstrand, 2019). The former is the ratio between the highest number of observations containing no HARP violation and the total number of violations. Specifically, it is useful to determine the set of observations that is strictly consistent with homothetic utility maximization. The latter emphasizes how much perturbation must be minimally made on each budget in order to remove all HARP violations overall. It thus provides extensive information on both the intensity and number of HARP violations for a given subject, though it is computationally more demanding.

B.1.1 Definitions of homothetic indices

B.1.1.1 The Homothetic Efficiency Index (HEI, Heufer, 2013)

Formally, let $\tilde{\mathbf{h}}_s = (h_s^i)_{i=1}^{30}$ a vector of numbers for subject s with $h_s^i \in [0, 1] \forall i$ the "homothetic efficiency level" associated with observation i . A "relaxed" version of

B.1.1.2 The Homothetic Houtman & Maks Index (H²MI, Heufer and Hjertstrand, 2015)

In order to compute the HMI, we follow Heufer and Hjertstrand (2015)'s suggestion to adapt the second algorithm proposed by Gross and Kaiser (1996) to HARP violations. In other words, we follow the exact same steps as for the computation of the HMI, except for replacing GARP with HARP (see the sub-sub-section A.2.1.2 from the appendix A).

B.1.1.3 The Homothetic Efficiency Vector (HEV, Heufer and Hjertstrand, 2019)

Formally, the HEV corresponds to the vector of homothetic efficiency levels h_s^i without the restriction that $h_s^i = h_s^* \forall i$. We compute the HEV for each subject by resorting to a MILP formulation as described in Heufer and Hjertstrand (2019).¹ Like them, we do not compute the exact HEV but a first-order Taylor approximation due to the problem's complexity.

Bulding upon the notion of \mathbf{h} -rationalization that they introduce, Heufer and Hjertstrand (2019) state the following theorem about the "relaxed" version of HARP:

- (i) there exists a well-behaved homothetic utility function $u(\cdot)$ that \mathbf{h} -rationalizes the set of observations: we have $u(\pi_s^i) \geq u(\pi_s^j)$ whenever $h_s^i p_s^i \pi_s^i \geq p_s^j \pi_s^j$.
- (ii) the set of observations satisfies HARP(\tilde{h}_s): we have $(p_s^i \pi_s^i)(p_s^j \pi_s^j) \geq h_s^i h_s^j$ for all observations i an j .
- (iii) there exists i numbers U_s^i such that $\forall i, j$:

$$U_s^i \leq U_s^j p_s^j \pi_s^j / h_s^j \tag{B.1}$$

Ideally, one would find the solution that minimizes the sum of inefficiencies, that is $\arg \min \{ \sum_{i=1}^n (1 - h_s^i) \}$. However, as emphasized by Heufer and Hjertstrand

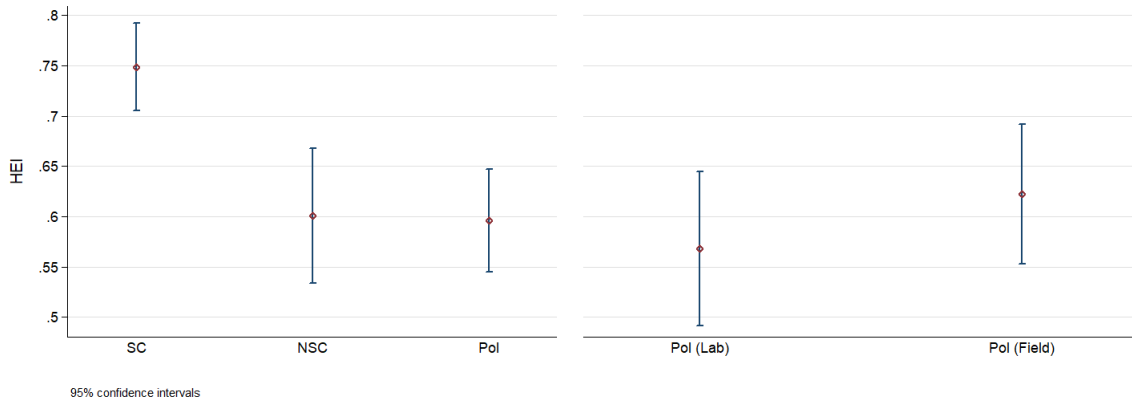
¹The advantages of relying on a MILP formulation to compute the HEV are the same as those exposed in the sub-sub-section A.2.1.3 from the appendix A regarding the computation of the Varian (1990, 1993)'s Efficiency Vector (VEV).

(2019), this problem is computationally demanding (if feasible). The authors propose to substitute the h_s^i with a first-order approximation: $\eta_s^i = \ln(h_s^i) \simeq -(1 - h_s^i)$. Such a formulation transforms the problem into $\arg \max \{\sum_{i=1}^n \eta_s^i\}$ subject to the log-linearization of equation (B.1): $\eta_s^j + u_s^i - u_s^j \leq \ln(\mathbf{p}_s^j \boldsymbol{\pi}_s^i)$, with $u_s^i = \ln(U_s^i) \forall i$. Together with the constraints $\eta_s^i \in (-\infty, 0]$ and $u_s^i \in (-\infty, +\infty)$, these equations constitute the MILP problem. Finally, the solutions \hat{h}_s^i are inferred from solutions $\hat{\eta}_s^i$ using $\hat{h}_s^i = \exp(\hat{\eta}_s^i)$.

B.1.2 Between-groups comparisons in homothetic indices

In this sub-section, we compare our experimental groups based on the three homothetic indices. The figure B.2 (resp. B.3) plots the mean HEI (resp. H^2MI) along with the 95% confidence interval for all the experimental groups. The table B.2 summarizes the information regarding the HEV. The table B.1 reports the p-values from several statistical tests across the groups.²

Figure B.2: Homothetic Efficiency Index (HEI) by groups



(a) Main groups

(b) Politicians' sub-groups

Note: Main groups = Politicians (Pol), Non-Students (NSC) and Students (SC). Politicians' sub-groups = Politicians from the Lab and Politicians from the Field.

All groups combined, the HEI averages 0.65 in our experiment. This implies that the budget lines need to be adjusted by 35% for the HARP violations of the average subject to be completely removed. We also observe marked heterogeneity across our groups in the figure B.2a. On average, Politicians (mean index of 0.596) generate an endowment waste of about 15 percentage points more than Students (0.749), but only 0.5 percentage points more than Non-students (0.601). The former difference is highly significant for all tests, whereas we systematically fail to reject the null

²In this table, we substitute the Kolmogorov-Smirnov test with the Epps-Singleton test in order to take into account the discrete nature of the H^2MI . The former is indeed not suitable for both continuous and discrete variables (and any mixture of the two), whereas the latter is (see Epps and Singleton, 1986).

of no difference between Politicians and Non-Students (table B.1). Moreover, in the figure B.2b, Politicians from the Field (0.568) appear a bit more consistent with homothetic utility maximization than Politicians from the Lab (0.622), but this difference is not statistically significant no matter the test under scrutiny. As a consequence, Politicians appear more distant from homothetic preferences than Students but we do not find evidence of a difference with a comparable control group or across Politicians' sub-groups.

Table B.1: p -values from statistical tests for differences in homothecity

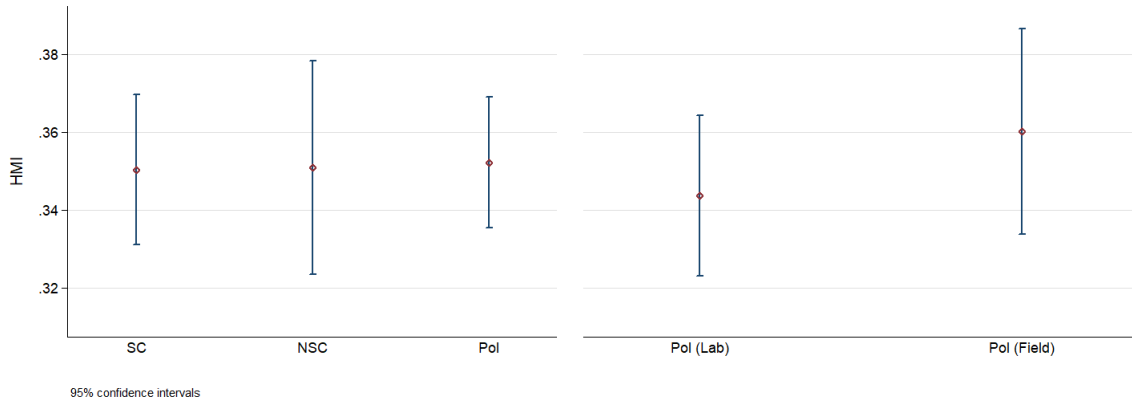
	Pol v. NSC			Pol v. SC			Lab v. Field		
	HEI	H ² MI	HEV	HEI	H ² MI	HEV	HEI	H ² MI	HEV
t-test	.919	.929	.835	<0.01	.884	<0.01	.293	.335	.291
FP test	.866	.464	.768	<0.01	.768	<0.01	.308	.677	.347
ES test	.875	<0.01	.904	<0.01	<0.01	<0.01	.258	.662	.124
Observations	230	230	229	262	262	261	139	139	138

Note: t-test: two-samples Student test. FP: Fligner-Policello. ES: Epps-Singleton.

HEV here stands for the minimum in the efficiency vector, applying suggestion by Varian (1993) to measuring homothecity. Similar results (unreported) are observed when comparing the average element in the efficiency vector.

Concerning the H²MI, it averages 0.351 in our experiment. This implies that between 19 and 20 allocations must be removed in order for the average subject to have no HARP violations. Yet, we do not observe the same heterogeneity across groups we previously found with the HEI. In particular, the average H²MI appears quite insensitive to the group under scrutiny: mean index of 0.352 for Politicians (0.344 for the Lab against 0.36 for the Field), 0.35 for Students and 0.351 for Non-Students. The pairwise differences regarding central tendencies (mean and median) are indeed not significantly different for all our groups (table B.1). However, the Epps-Singleton tests on differences in entire distributions yield significant differences between Politicians and both our control groups. Looking at the distributions, it appears that the distribution of the H²MI for both control groups is somewhat less concentrated than the distribution for Politicians. For instance, there are about 71% of the Politicians

Figure B.3: Homothetic Houtman-Maks Index (H²MI) by groups



(a) Main groups

(b) Politicians' sub-groups

who need between 18 and 24 removals to be HARP-consistent, against 58% for Non-Students and 59% for Students. At the same time, the difference is not significant between Politicians' sub-groups.

We summarize all these additional results with respect to between-groups comparisons in closeness to homothetic preferences as:

Additional result 1 (homothetic preferences among Politicians' sub-groups):

Both Politicians' sub-groups do not differ in their proximity to homothetic preferences.

Additional result 2 (homothetic preferences in Politicians v. Non-Students):

Politicians do not differ from Non-Students in their proximity to homothetic preferences.

Additional result 3 (homothetic preferences in Politicians v. Students):

Politicians do not violate HARP more often than Students, but the extent of their violations is greater.

So far, the evidence suggests that not accounting for the specific additional violations due to the homothetic preferences assumption is likely to entail increasing noise when estimating a homothetic utility function (like a CES), as compared to only assuming consistent preferences. But such a statement does not mean that estimating a CES utility function is not useful. Indeed, the results on the HEV

shows that most subjects ultimately fall close to homothetic utility maximization: despite few big violations, most efficiency levels in the vector are very close to one (see summary statistics in table B.2).³ The average required adjustment is pretty low as only 4% of the budget is wasted on average across all decisions made by our subjects: observed decisions are overall 96% as efficient as maximizing homothetic utility. Furthermore, the average mean of homothetic efficiency levels in the HEV is above 0.95 for all groups. The same applies to the average first quartile of homothetic efficiency levels: for all groups, more than 3 out of 4 observations within a subject's dataset are associated with violations less than 5% of the endowment on average. In other words, sticking to the CES utility function for the many subjects who display only marginal violations of homothecity still sounds reasonable.

Interestingly, the results about homothetic utility maximization are somewhat "sharper" than the results about utility maximization without any additional restriction on functional form, as exposed in the section 3 from the chapter 1.⁴ Indeed, the average welfare loss one needs to acknowledge from assuming utility maximization is 25 percentage points less (average HEI of 0.65) than the welfare loss implied by assuming homothetic utility maximization (average AEI of 0.89). Furthermore, abiding by HARP requires removing 17 observations more than abiding by GARP on average. Additionally, we do not find any individual dataset strictly compatible with HARP, whereas GARP-consistent subjects represent about a third of our sample. When comparing vector of efficiencies, the main result is that assuming homothecity implies that most observations need an adjustment, whereas most observations entail no welfare loss if we consider only consistent preferences. The homothetic preferences assumption is therefore much more demanding than assuming that behavior is coherent with maximizing some utility function. In other words, the utility maximization model alone may perform better in our dataset than when homothecity is added, in the sense that it amounts to sacrificing less money. Similarly, the statistically-significant differences we observe across our groups are bigger, whereas the non-significant differences are smaller. We interpret such finding as

³As concerns between-groups heterogeneity, the results on the minimum in the HEV are very similar to the results from the HEI and the statistical tests all yield similar conclusions (table B.1).

⁴The three homothetic indices are basically the homothetic versions of the three rationality indices which the chapter 1 is built upon, so that they are readily comparable.

Table B.2: HEV summary across groups

(a) Mean summary						
	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
Mean	0.949	0.956	0.953	0.952	0.974	0.960
Minimum	0.606	0.659	0.633	0.642	0.773	0.684
First Quartile	0.949	0.954	0.951	0.953	0.974	0.960
Median	0.988	0.989	0.989	0.988	0.995	0.991
Third Quartile	0.998	0.998	0.998	0.998	0.999	0.998
Maximum	1.000	1.000	1.000	1.000	1.000	1.000
(b) Median summary						
	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
Mean	0.962	0.971	0.966	0.974	0.987	0.976
Minimum	0.590	0.704	0.677	0.674	0.832	0.732
First Quartile	0.979	0.977	0.978	0.983	0.993	0.986
Median	0.997	0.998	0.998	0.998	0.999	0.998
Third Quartile	0.999	1.000	1.000	1.000	1.000	1.000
Maximum	1.000	1.000	1.000	1.000	1.000	1.000

Note: The values for each group correspond to the average or median values of each considered statistic computed at the individual level. For instance, the minimum values rely on computing $\min\{h_s^1, \dots, h_s^{30}\}$ for each subject s and averaging these values over all subjects in the corresponding group.

indicative that testing for homothetic preferences is in a sense more "revealing" of group differences than testing for choices' consistency.

We also note that our results are fairly compatible with existing findings. For instance, Heufer (2013) emphasizes an the average HEI of 0.705 with the student sample from Fisman et al. (2007), which is close to our Students' average HEI (0.749).⁵ Similarly, he does not find any HARP-consistent subject. Summary statistics on the

⁵That in our experiment we obtain slightly higher levels of HEI is not surprising considering that we have a lower number of decisions and that the HEI can only decrease when the number of decisions increases.

HEV for Students in table B.2a are also strikingly close to the corresponding table in Heufer and Hjertstrand (2019).

B.1.3 Closeness to homothetic preferences: power computations

In addition, we compare our subjects' closeness to homothetic preferences to a counterfactual simulated dataset composed of uniform-random decision-makers. Indeed, the homothetic indices do not indicate whether violations of HARP are likely to be detected or not and may consequently suffer from a lack of power. We assess their ability to rule out irrational behavior by comparing the indices of our actual subjects with indices obtained from Monte-Carlo simulations. Specifically, for each subject we generate a hundred simulated subjects choosing uniformly over the set of possible allocations for the 30 lines that the subject actually faced. Then we compute the HEI, the H²MI and the HEV for each simulated subject and obtain a simulated distribution of the indices under random (irrational) choice. Ultimately we subtract the average value of each index to the index of the actual subject to obtain a measure of the predictive success of the homothetic utility maximization model, that is its ability to make sense of observed data as opposed to an alternative random choice model.⁶ Such a measure has been called the "Selten score" (Dean and Martin, 2016) and is directly inspired by the predictive power measure from Beatty and Crawford (2011). The table B.3 summarizes our results.

Overall, our subjects fall much closer to homothetic utility maximization than their uniform-random counterparts. In particular, the former outperform the latter by more than 35 percentage points for the HEI and by more than 28 percentage points for the minimum in the HEV on average. Although we observe some differences between the experimental groups, the magnitude of the difference between actual and simulated subjects remains economically significant for all groups. Besides, the average score is generally significantly different from 0 no matter the group (t-test: $p < 0.01$ whenever significant). Only for the H²MI do the tests fail to reject

⁶The uniform benchmark may not be the best counterfactual but it naturally applies to any choice environment and is heavily relied on within the literature on revealed preference. See Bronars (1987) for a discussion.

Table B.3: Summary statistics for Selten scores on each homothetic index across groups

	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
HEI	0.271 (0.308)	0.320 (0.296)	0.296 (0.302)	0.304 (0.325)	0.449 (0.246)	0.351 (0.298)
H ² MI	0.001 (0.086)	0.016 (0.110)	0.009 (0.099)	0.006 (0.131)	0.004 (0.109)	0.007 (0.111)
HEV (min)	0.208 (0.297)	0.259 (0.283)	0.234 (0.290)	0.242 (0.302)	0.379 (0.225)	0.287 (0.280)
HEV (mean)	0.074 (0.046)	0.079 (0.046)	0.076 (0.046)	0.076 (0.053)	0.097 (0.030)	0.084 (0.044)

Note: The values for each group and each homothetic index correspond to the mean and standard deviation (in parentheses) of the Selten score for the corresponding homothetic index, computed as the difference between the actual index and the mean index of a hundred simulated uniform-random subjects.

The "HEV (min)" (resp. "HEV (mean)") row exposes results for the minimum (resp. average) in the vector.

the null hypothesis of no difference between actual subjects and simulated ones. Such a result means that on average subjects do not deviate less often from HARP than what would random players do, but they manage to deviate most of the time only by small amounts whereas random players generally fail the test by far.

Interestingly, comparing these results on homothetic indices with the results obtained on rationality indices in the chapter 1, we can observe that the Selten scores are much higher for the former than for the latter (except for the H²MI). In other words, if adding the homothecity restriction on preferences is more demanding than simply assuming consistent preferences, it is actually even harder on random choosers. Assuming homothetic preferences thus comes with a higher discriminatory power as compared to the simple utility maximization model.

B.2 Complements on the empirical strategy

B.2.1 Description of the structural estimation process

From the resolution of the consumer program with a CES utility function, we derived the following individual-level econometric specification for subject s and modified dictator game i :

$$p_{s,d}^i \pi_{s,d}^i = \frac{\gamma_s}{(p_{s,d/r}^i)^{\kappa_s} + \gamma_s} + \epsilon_s^i \quad (\text{B.2})$$

with the error term ϵ_s^i assumed to be independent and identically normally distributed with mean zero and variance σ_s^2 .

We estimate (B.2) through a maximum likelihood procedure. Given that our data is censored from below ($p_{s,d}^i \pi_{s,d}^i \geq 0$) and from above ($p_{s,d}^i \pi_{s,d}^i \leq 1$), the econometric specification translates into the following likelihood function:

$$\begin{aligned} \mathcal{L}(\gamma_s, \kappa_s, \sigma_s) = & \prod_{i: p_{d,s}^i \pi_{d,s}^i = 0} \left[\Phi \left(-\frac{\gamma_s}{\sigma_s ((p_{s,d/r}^i)^{\kappa_s} + \gamma_s)} \right) \right] \\ & \times \prod_{i: 0 < p_{d,s}^i \pi_{d,s}^i < 1} \left[\frac{1}{\sigma_s} \phi \left(\frac{1}{\sigma_s} \left(p_{d,s}^i \pi_{d,s}^i - \frac{\gamma_s}{(p_{s,d/r}^i)^{\kappa_s} + \gamma_s} \right) \right) \right] \\ & \times \prod_{i: p_{d,s}^i \pi_{d,s}^i = 1} \left[1 - \Phi \left(\frac{(p_{s,d/r}^i)^{\kappa_s}}{\sigma_s ((p_{s,d/r}^i)^{\kappa_s} + \gamma_s)} \right) \right] \end{aligned}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function and $\phi(\cdot)$ is the normal density function.

We implement the two-stages procedure of Fisman et al. (2007) in order to run the estimation: first we estimate (B.2) by non-linear-least squares, and second we use the estimated parameters from this first stage as initial values for the Nelder-Mead search algorithm which solves the maximum likelihood problem. Within the first stage, we select $\alpha_s = 0.5$ (i.e., $\gamma = 1$) and $\rho = 0$ (i.e., $\kappa = 0$) as initial values, which correspond to the mid-points of each corresponding tradeoff. Such a process provides individual estimates of γ_s and κ_s . We then use such estimates to recover estimates

of underlying CES parameters α_s and ρ_s from $\hat{\alpha}_s = \left(\hat{\gamma}_s^{1/(1+\hat{\kappa}_s)} \right) / \left(1 + \hat{\gamma}_s^{1/(1+\hat{\kappa}_s)} \right)$ and $\hat{\rho}_s = \hat{\kappa}_s / (1 + \hat{\kappa}_s)$.

Finally, we are also confronted to two computational problems when running the estimation. Though rather technical, such problems may have had a strong impact on our findings if neglected.

The first problem deals with the fact that the algorithm we resort to may not converge towards solutions that are restricted to the theoretical parameters' domain. In particular, we uncover 22 subjects for which $\hat{\rho}_s > 1$ (13 Politicians, 7 Non-students and 2 Students) and 1 subject for which $\hat{\alpha}_s < 0$ (a Non-student) when running the above-described estimation process. Such a situation may happen due to both the fact that the algorithm can converge towards a local solution (instead of a global one) and to the fact that the above likelihood function does not incorporate the theoretical constraints we set on the parameters. In other words, the algorithm is unaware that it should search for solutions such that $\hat{\rho}_s < 1$ and $\hat{\alpha}_s \in [0, 1]$ and may find some local ones violating these constraints. We propose fixing this issue through a change of variable and estimate $K_s = \ln(\kappa_s + 1)$ and $\Gamma = \ln(\gamma_s)$ instead of κ_s and γ directly. The first change restricts κ_s to be above -1, hence ρ_s to be below 1. The second change restricts γ_s to be above 0, hence α_s to be bounded between 0 and 1. Then it is straightforward to compute $\hat{\kappa}_s$ (hence $\hat{\rho}_s$) and $\hat{\gamma}_s$ (hence $\hat{\alpha}_s$) from \hat{K}_s and $\hat{\Gamma}_s$.

The second problem is related to the way data analysis softwares (and currently Stata) format the variables they store: they typically cannot deal with "infinity" and thus come with some pre-defined arbitrary boundaries for what is a really small/big number. In other words, the software will approximate at some point the true number due to limited storage capacities. More specifically, if \hat{K}_s happens to be very negative (i.e., the subject is very sensitive to equality considerations), the software will store $\exp(\hat{K}_s)$ as 0 and thus compute $\hat{\kappa}_s = -1$, which results in a missing value for $\hat{\rho}_s$. Our fix consists in setting a scalar $\bar{\kappa}_s$ such that $\hat{\kappa}_s = \bar{\kappa}_s - 1$ whenever $\hat{K}_s \leq \ln(\bar{\kappa}_s)$.⁷ In other words, for these "numerically constrained" subjects, we compute an upper bound on $\hat{\rho}_s$. As a consequence, all these subjects will have the same $\hat{\rho}_s = (\bar{\kappa}_s - 1) / \bar{\kappa}_s$, so that the estimation process loses some precision.

⁷We (arbitrarily) implement $\bar{\kappa}_s = 10e^{-9}$, which is the smallest integer power of ten one can use as an input value to a logarithmic function in Stata.

Furthermore, the same kind of "numerical constraint" applies to $\hat{\alpha}_s$. In particular when $\hat{\kappa}_s$ is really close to -1, $\hat{\gamma}_s^{1/(1+\hat{\kappa}_s)}$ will rapidly get too big to be properly stored if $\hat{\gamma}_s$ is above 1. As a consequence, we approximate the value by setting $\hat{\alpha}_s = 1$ for such a case (thus considering that when big numbers are involved, adding 1 is negligible, hence dividing the number by itself yields 1).

B.2.2 Description of the estimation restrictions

In this sub-section, we detail how we define the two subjects-based restrictions and the two observations-based restrictions that we implement prior to running the above structural estimation. These restrictions aim at taking into account important departures from homothetic utility maximization that may lead to a noisy measure of underlying preferences. The problem is therefore to define what "important" means. On the one hand, the restrictions should allow for some deviations because it is realistic (and empirically sound) to assume that subjects may differ from homothecity (e.g., due to fatigue, miss-clicks, or other small errors). On the other hand, allowing for too big deviations bears the risk that the revealed preference tests are not meaningful anymore, which is a problem of power.

B.2.2.1 Subjects-based restrictions

For the subjects-based restrictions, we select the threshold which separates "homothetic-enough" subjects from the remainder of subjects based on power computations. Such analyses precisely provide a way of assessing how much one can afford to relax the revealed preference conditions without hinging too much on the experiment's ability to reveal a violation of the conditions, hence the optimal score for a given homothetic index.⁸

Restriction on the HEI: $HEI \geq 0.51$.

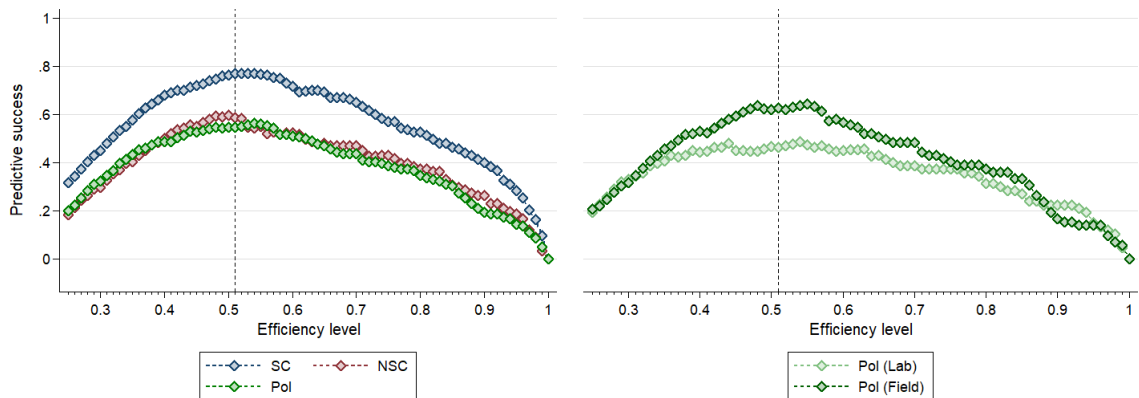
To find the optimal score for the HEI, we rely on an individual measure combining efficiency and power advocated by Beatty and Crawford (2011). The procedure is exactly the same as the one described in the sub-section 6.1 from the chapter 1, except it is adapted for the HARP(\tilde{h}_s^*) test (in lieu of the GARP(\tilde{e}_s^*) test). Specifically, we compute the predictive success measure for subject s over several efficiency levels \tilde{h}_s^* as follows: i) check whether or not subject s complies with HARP(\tilde{h}_s^*) (binary pass rate), ii) simulate a hundred uniform subjects choosing at random on the same thirty budget lines as subject s ⁹, iii) compute the proportion

⁸For an insightful discussion and a presentation of several ways to adress the power issue of an experimental testing revealed preference conditions, see Andreoni et al. (2013).

⁹In practice, the simulated choices that we use to compute the HEV are the same as the ones that we used to compute the VEV.

of simulated subjects that do not comply with HARP(\tilde{h}_g^*) (power). Repeating such procedure for various efficiency levels given an arbitrarily fine grid provides a way to identify the efficiency level which maximizes the predictive power (hence a good candidate for defining "homothetic-enough" behavior). The figure B.4 summarizes the results on predictive success: for each efficiency level, the graph plots the average predictive success within the corresponding group.

Figure B.4: Mean predictive success for different efficiency levels across groups (HEI)



(a) Main groups

(b) Politicians' sub-groups

Note: We compute the individual predictive success for a given efficiency level as the difference between whether or not the subject passes the HARP test at such efficiency level and the probability that uniform-random subjects pass the same test. Each dot plots the average predictive success for the corresponding efficiency level within the group.

We find that the maximal predictive success is about 0.64, reached for an efficiency level of 0.51. In other words, our average subject is passing the HARP(**0.51**) test when only 36% of simulated subjects achieve the same result. Consequently, the homothetic utility maximization hypothesis that allows for a waste amounting to approximately half the budget outperforms by far an alternative random behavior hypothesis. We also observe that for all efficiency levels ranging from 0.37 to 0.7, our subjects fare better with respect to homothecity than at least half the simulated subjects. Ultimately, our subjects do deviate from HARP rather substantially, but such violations do not come close to the violations entailed by random choosers. We will thus select $HEI = 0.51$ as the threshold that identifies subjects sufficiently close

to homothetic preferences and for which we will run the estimation. Such a restriction implies that only 247 out of the 353 subjects (69.97%) are kept in the analysis (85 Politicians (61.15%), 103 Students (83.74%) and 59 Non-Students (64.84%)).

We also recover previous results with respect to between-groups comparisons. First, the homothetic utility maximization model fits Students' data the best: maximal average predictive success of about 0.77, reached for an efficiency level of 0.53. Second, the pattern of predictive success over efficiency levels for Politicians is very similar to the one for Non-Students, which means that there is no conspicuous difference between both groups regarding the optimal efficiency-power tradeoff. Third, we observe a disconnection in predictive success between Politicians from the Lab and Politicians from the Field for intermediate efficiency levels (between 0.4 and 0.7): the latter appear closer to homothetic utility maximization than the former. The curve for Politicians from the Lab is indeed much smoother with no real peak at some efficiency level.

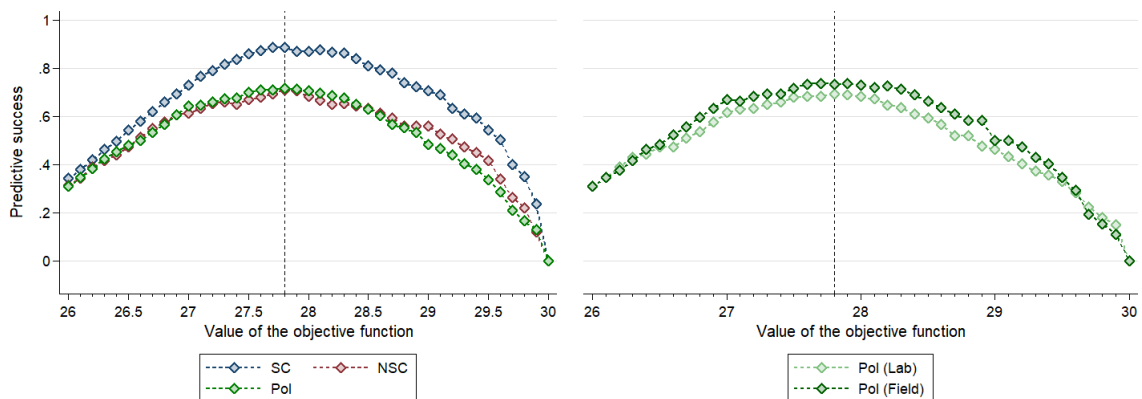
Restriction on the average in the HEV: average HEV ≥ 0.927 .

Finding the optimal score on the average element in the HEV involves a slightly different reasoning. This is because comparing vectors is not as straightforward as comparing numbers. With the HEI, we can immediately compare actual subjects with simulated ones based on whether or not they pass the HARP(\tilde{h}_g^*) test at some efficiency level \tilde{h}_g^* . However, if we allow efficiency levels to differ within the efficiency vector, then it is likely that we observe simulated subjects failing at the test that actual subjects pass even though they are as close to homothecity as the latter. Consider for instance the case where an actual subject only needs a 10% shift on the first budget line to comply with HARP, due to a cycle involving both the first and the second observations. Based on the same budget lines, a simulated subject may only need the same 10% shift on the third budget line, due to a cycle involving both the third and the fourth observations. Surely the two subjects seem as close to homothecity: they simply made different mistakes of the same magnitude. However, they are both failing at the test that the other is passing. Consequently, we cannot directly test whether simulated subjects fare better or worse than actual subjects based on the adjustments computed in the latter's efficiency vectors.

What we can do is compare what subjects are wasting overall, which is the

objective function of the MILP procedure for computing the HEV. In other words, we state that an actual subject is closer to homothecity than a simulated subject if the sum of efficiency levels in the HEV is bigger for the former than for the latter (hence a lower waste). We therefore adapt the predictive success measure for subject s over several values of the objective function as follows: i) check whether or not the sum of efficiencies from the HEV of subject s is above the current value, ii) simulate a hundred uniform subjects choosing at random on the same thirty budget lines as subject s , iii) compute the proportion of simulated subjects whose sum of efficiencies from their HEV is below the current value. Repeat the procedure for various values of the objective function from 0 to 30 given an arbitrarily fine grid. The figure B.5 summarizes the results: for each value of the objective function, the graph plots the average predictive success within the corresponding group.

Figure B.5: Mean predictive success for different objective values across groups (HEV)



(a) Main groups

(b) Politicians' sub-groups

Note: We compute the individual predictive success for a given value of the objective function as the difference between whether or not the subject is wasting more money than such value overall and the probability that uniform-random subjects waste the same amount. Each dot plots the average predictive success for the corresponding efficiency level within the group.

We find that the maximal predictive success is about 0.77, reached for a sum of efficiencies of 27.8. In other words, our average subject is capable of reaching 92.7% of the overall target defined by homothetic preferences, whereas only 23% of

simulated subjects are achieving the same result. Such a value translates into an average efficiency level of about .927, which will we use as a threshold on the average element in the HEV to identify subjects close-enough to homothetic preferences. Such a restriction implies that only 287 out of the 353 subjects (81.3%) are kept in the analysis (105 Politicians (75.54%), 114 Students (92.68%) and 68 Non-Students (74.73%)).

Again, we also observe that the homothetic utility model is best describing Students' data (with a maximal predictive success of about 0.89), whereas it fits Politicians and Non-Students' data equally well. Interestingly, the difference between Politicians from the Lab and Politicians from the Field is much less conspicuous than it was with the predictive success computed for the HEI. We interpret such finding as indicative that the difference between the two groups is mostly driven by Politicians from the Lab violating HARP in a larger extent for the biggest violation. However, the two groups only marginally differ in terms of the less-extreme violations. Such a statement was already readable in the table B.2: the difference in the minimum in the HEV between the two groups is much bigger than the difference in any other statistic (mean or quartiles).

B.2.2.2 Observations-based restrictions

Subjects-based restrictions are interesting insofar as they allow for the same analysis to be run for all subjects. The homothetic status of each subject is defined based on his behavior in the entire game and all decisions are used in the structural estimation process. However, they entail losing from about a quarter to about a third of the whole sample. Considering our experiment targets a hard-to-recruit population, the power issue implied by such a loss is particularly problematic. Within the observations left aside, it is likely that some convey useful information. It may even be the case that some subjects are discarded because of few observations that make them inconsistent with homothetic preferences. In such cases, the useful information is predominant. At the same time, some subjects that are kept may violate the axioms many times by small amounts. Furthermore, even though the selected thresholds for subjects-based restrictions are scientifically-grounded (maximizing the predictive success), it is noticeable that a much wider range of efficiency levels/values of the objective function corresponds to rather high levels of predictive success (though

not maximal). It is thus likely that subjects-based restrictions are somewhat too restrictive.

Based on such observations, we propose to take into account violations of homothecity by withdrawing observations instead of subjects. In other words, in addition to a strategy that keeps all observations but drops some subjects, we consider a complementary strategy that keeps all subjects but drops some observations.¹⁰

Restriction on the H²MI

For this restriction, we consider the conservative stance according to which all observations not complying with HARP are removed for each subject.¹¹ This seems fairly demanding as it entails keeping only a third of observations per subject on average. In a sense it is even more demanding than subjects-based restrictions considering that a lot more observations are neglected in such a case.

Restriction on the HEV

In order to take into account that removing all observations involved in a small violation is likely too conservative, we define this second restriction based on all observations which are responsible for at most 5% budget waste. For a given subject, any observation which is associated with an efficiency level below 0.95 is discarded. Although it may echoe the 5% golden standard from statistical analysis, such a number is obviously arbitrary. However, it appears appealing as it corresponds to a small budget waste but also to the very large majority of observations. Indeed, there are only about 16.13% observations that fall below this threshold. Furthermore, if we allow for a 5% average adjustment in the HEV, hence a sum of efficiencies of 28.5, the average subject is still exposed to a highly-restrictive test in the figure B.5.

¹⁰Although the literature this chapter borrows from methodologically focuses on subjects-based restrictions, withdrawing observations instead of subjects is nothing new in experimental economics.

¹¹Such a procedure was directly inspired by the following sentence in the introduction of Heufer and Hjertstrand (2015): "The [Houtman and Maks (1985)'s] method has the additional advantage that researchers can restrict further analysis of the data set to this maximal subset."

B.3 Complementary analyses on group comparisons from raw data

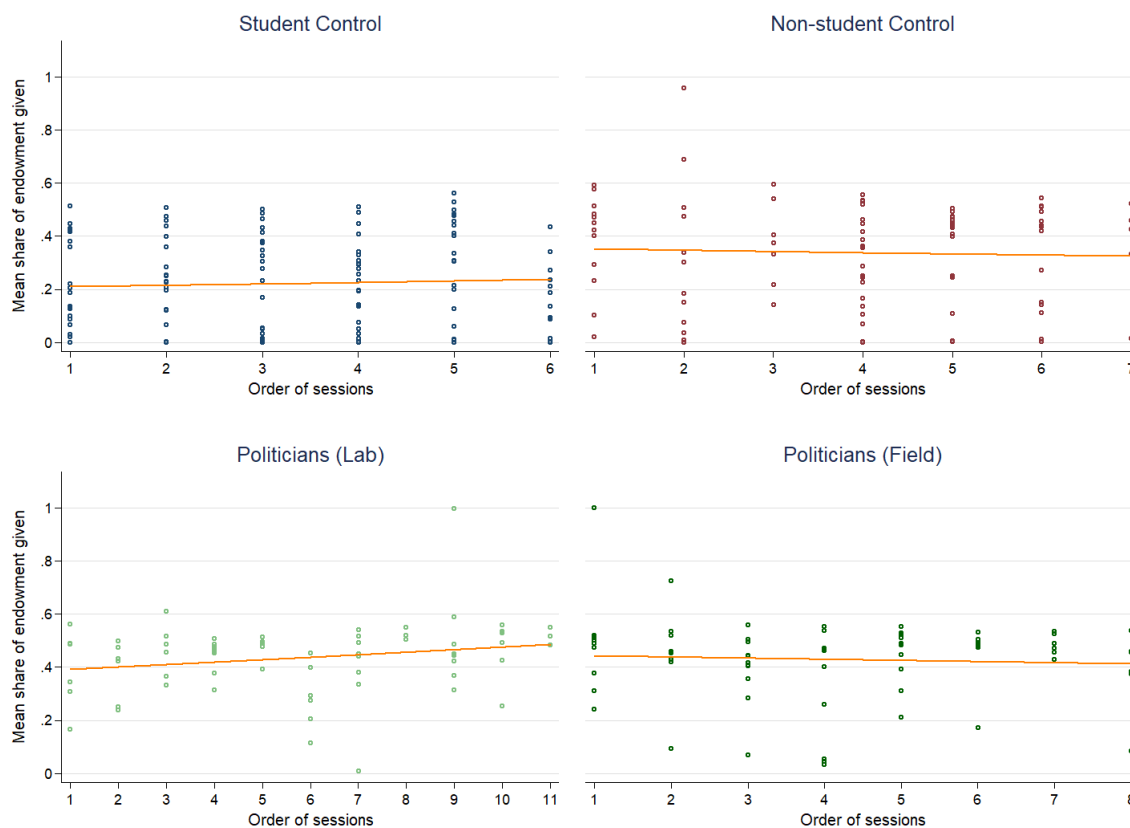
B.3.1 Robustness check on cross-sessions contagion effect

One concern with our experimental procedures is the possible risk of cross-sessions contagion for Politicians from the Lab and Non-Students. Indeed, the recruitment process partly relied on a word-to-mouth communication between past subjects and potential future subjects for both these groups. In particular, Politicians from the Lab were kindly asked to broadcast our participation call within their networks.¹² Though in both case subjects signed a non-disclosure form and were orally reminded the importance of avoiding any bias for the results to be meaningful, there is no guarantee that subjects did abide by the rules. As a consequence, it is possible that subjects from late sessions had access to uncontrolled prior information about the content of the experiment before entering the laboratory as compared to subjects from early sessions, decreasing the comparability across sessions. It is also possible that such a recruitment channel only worked with subjects (either past or potential) who happen to have preferences weighted in some way. In either case, we could be confronted to a contagion effect across sessions that may affect our results. In order to quantify such effect, we look at the relationship between the ordering of sessions and the average share of endowment given (figure B.6).

For all groups but Politicians from the Lab, there is no economically nor statistically meaningful relationship. The predicted values from a regression of session's rank on the average share of endowment given (the orange lines) indeed form a flat line. On the contrary, the group which bears the highest risk of cross-sessions contagion (Politicians from the Lab) does show some positive relationship: Pearson's correlation coefficient of 0.2058, which is weakly statistically significant ($p = 0.0948$). Nevertheless, the magnitude and the level of significance of such a correlation are rather low, which prompts us into thinking that it is actually not an important

¹²On the contrary, such a contagion effect is unlikely for both Students and Politicians from the Field. The former were invited at random from the ORSEE database with not prior information about the experiment they were invited to. The latter participated in sessions organized in different towns.

Figure B.6: Sessions' rank and average share of endowment allocated to the recipient



Note: Each dot corresponds to the share of endowment given by a subject on average across all decisions.

Orange lines fit a linear regression model of session's rank on average share of endowment given.

confound. Specifically, observed behavior across Politicians' sub-samples do not differ significantly and, since behavior is rather homogeneous across sessions in both groups, our main comparisons with both control groups are likely to stand. Furthermore, when running the same analysis on the average fraction of tokens allocated to the recipient, we do find the same pattern but the relationship is not significant anymore. Besides, we also looked at the correlations between sessions' rank and the number of times the subject is consistent with either full equality or full efficiency. Neither of the correlation coefficients are significant for any group. Finally, the effect of another variable is likely to affect the relationship we found between average share of endowment given and sessions' rank: the number of subjects per session. Indeed, some sessions happened with only 3 subjects, which is a partic-

ularly low number (the tradeoff being maintaining the session or losing all three observations). Though subjects were still previously unacquainted (we made sure of that prior to deciding upon maintaining the session or not), such a feature may have distorted behavior since it eased the identification of the potential partner in the game. Hence a possible interplay of partner's personal characteristics with the decision.¹³ Another interpretation is that sessions with a low number of participants somehow weaken the anonymity conditions between subjects since subjects basically know that there is a high possibility that their partner correctly guess who they are, which could trigger an increase in giving.¹⁴ Interestingly, both sessions with only 3 subjects are the ones with the highest mean values for the average share of endowment given. This could affect the present analysis since these sessions occurred near the end of the first experimental period. Indeed, upon withdrawing them, the

¹³For example, consider gender as an obvious characteristic a subject can infer from others in the room. Some studies show that knowing the partner's gender may affect the decision. For instance Saad and Gill (2001) find that both men and women give more in standard Dictator Game to women than to men, while on the contrary women are less prone to give to female partners than to male or unknown partners based on the same game in Ben-Ner et al. (2004). Similarly, the gender composition of a group also matters: groups where women are in majority are more generous and achieves a higher degree of equality in Dufwenberg and Muren (2006). Yet they authors also note that the groups with two men and one women are the most generous. Such a pattern is quite interesting since both our sessions with 3 subjects are not composed of same-sex subjects (the two remaining possible configurations occur). Of course, gender is not necessarily the only characteristic that may matter: our point is that, with very few subjects in the room, potential partners' personal characteristics may come in the way.

¹⁴There are indeed numerous studies emphasizing that people dislike being perceived as selfish or "unfair". For instance Dana et al. (2007) show that the degree of transparency of the dictator game is negatively related to selfishness: when there is a higher probability to "hide" one's action, the dictator tend to give less to the recipient. Such a relationship is observed when the dictator is given the possibility to choose to hide her action or when the uncertainty is driven by chance. Similarly, in a similar dictator game with publicly disclosed decisions and uncertain implementation of the dictator's choice, Andreoni and Bernheim (2009) test the hypothesis that "people like to be perceived as fair" (in the sense of the 50-50 norm). Indeed, they find that most subjects target fairness when the probability of the decision to be randomly determined is low and that there is a sharp distancing from fairness in favor of the randomly determined allocation when such a probability increases. This significant negative effect of decision uncertainty on giving in the dictator game is also reported in Engel (2011)'s meta-analysis.

Pearson's correlation coefficient is reduced by approximately a third (0.1444) and is not statistically significant anymore ($p = 0.2664$).¹⁵

¹⁵However, such an effect of the number of subjects within each session on subjects' behavior is not likely to be a concern when the number increases. Indeed when very few people are in the same room, a subject can easily react to his potential partner's personal characteristics. This is not the case when there are more people since the information about any potential partner gets more diffuse. Especially since we did not lay the emphasis on any of these characteristics. Indeed, correlations between behavior in the game and the number of subjects in the session are never economically or statistically significant no matter the group under scrutiny (even when pooling the Politicians altogether) or the variable we use for "behavior" (average fraction of tokens are average share of endowment given). Of course, we may lack some variation in the number of subjects per session or in the number of sessions to actually detect an existing effect (although the Non-students show a very diverse profile on that respect), but in any case we believe it would not be a major driver of our results.

B.3.2 Power of the statistical tests comparing Field and Lab on raw data

In this sub-section, we run Monte-Carlo simulations to assess the power of the statistical tests we implement to detect a difference across Politicians' sub-groups in average pass rate/fraction. The tests indeed fail to reject the null of no difference. The simulation procedure borrows from Müller (2019) and is akin to the one used in the section 6.2 from the chapter 1. Given a behavioral measure computed for one baseline group (b) and one "treated" group (t) with sample sizes n_k , empirical means μ_k and empirical standard deviations σ_k ($k \in b, t$), we repeat a hundred times the following steps for each "treatment effect" $d \in \{0, 0.025, 0.05, 0.075, 0.1, 0.125\}$:

1. Draw n_b iid observations from a normal distribution with mean $\hat{\mu}_b$ and standard deviation $\hat{\sigma}_b$.
2. Draw n_t iid observations from a normal distribution with mean $\hat{\mu}_t \pm d$ and standard deviation $\hat{\sigma}_t$ (the sign of d depends on the direction of the empirical difference between the groups).
3. Replace draws lying outside of the behavioral measure's boundaries with the corresponding boundary (e.g., the endowment share is necessarily $\in (0, 1)$ so that any draw strictly above 1 (resp. below 0) should be restricted to 1 (resp. 0)).
4. Store the p-values from applying the statistical tests to detect differences between the two simulated distributions.

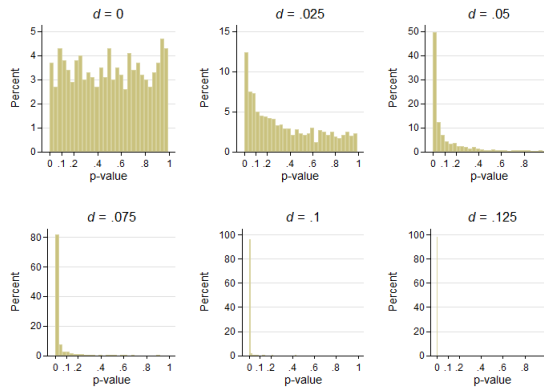
Distributions of p-values from all tests on the differences in average pass fraction (resp. pass rate) between our Politicians' sub-groups are depicted in the figure B.7 (resp. B.8). The results are also summarized in table B.4.

Overall, we find that our tests reach usual thresholds for type-II errors (power of 80% or 90%) under usual 5% or 10% thresholds for type-I errors when the difference between the two groups is at least 7.5 percentage points. For instance, 85.7% (resp. 91%) of the Fligner-Policello tests on average pass fraction yield a p-value below 0.05 (resp. 0.1) for $d = 0.75$ between Politicians from the Lab and Politicians from the Field. In other words, our experiment can convincingly rule out any mean difference in average pass fraction/rate bigger than 7.5 percentage points between

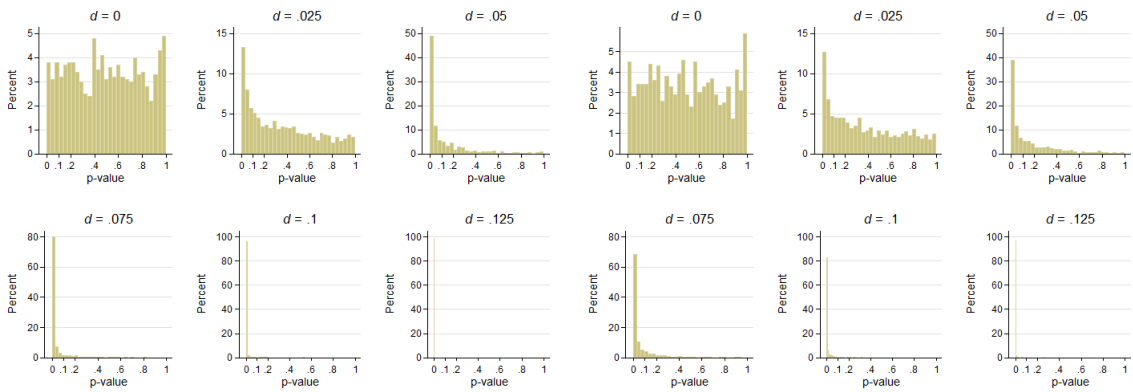
APPENDIX B

the two Politicians' sub-groups. Furthermore, our tests also have a decent chance to detect a 5 percentage points difference, although the odds for the same Fligner-Policello test fall to 56.1% (resp. 65.6%). We conclude from such exercise that the differences between Politicians from the Lab and Politicians from the Field are likely limited, since we would have been able to detect even moderately bigger differences. In particular, we would have been able to detect smaller differences than the one we observe between Politicians and Non-Students (about 10 percentage points).

Figure B.7: P-values from simulations of a hundred tests on differences in average pass fraction



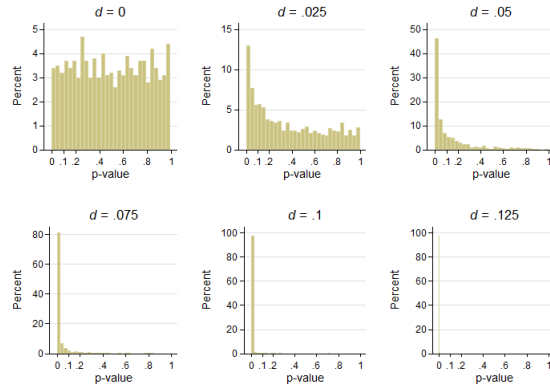
(a) t-test



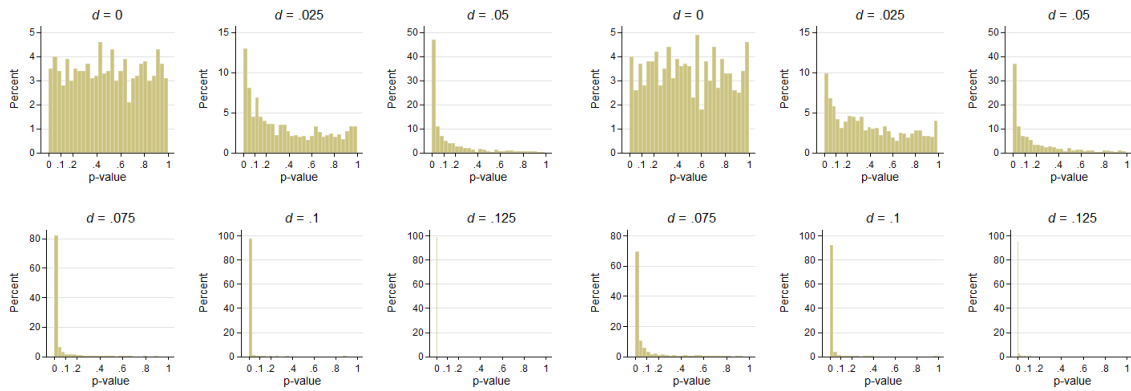
(b) Fligner-Policello test

(c) Kolmogorov-Smirnov test

Figure B.8: P-values from simulations of a hundred tests on differences in average pass rate



(a) t-test



(b) Fligner-Policello test

(c) Kolmogorov-Smirnov test

Table B.4: Summary statistics for statistical tests based on simulated data

		$d = 0$		$d = 0.025$		$d = 0.05$		$d = 0.075$		$d = 0.1$		$d = 0.125$	
		Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Average pass fraction	t-test	0.50	0.29	0.36	0.30	0.13	0.21	0.03	0.08	0.00	0.02	0.00	0.00
	FP test	0.50	0.29	0.36	0.30	0.14	0.21	0.03	0.08	0.00	0.02	0.00	0.01
	KS test	0.49	0.29	0.38	0.30	0.17	0.22	0.06	0.11	0.01	0.03	0.00	0.01
Average pass rate	t-test	0.50	0.29	0.37	0.31	0.14	0.21	0.03	0.08	0.00	0.03	0.00	0.00
	FP test	0.50	0.29	0.38	0.32	0.14	0.21	0.03	0.09	0.00	0.03	0.00	0.01
	KS test	0.50	0.29	0.40	0.30	0.18	0.23	0.06	0.13	0.01	0.05	0.00	0.01

B.4 Complementary analyses on group comparisons from parameters

B.4.1 Distributions of estimated parameters for each restriction strategy

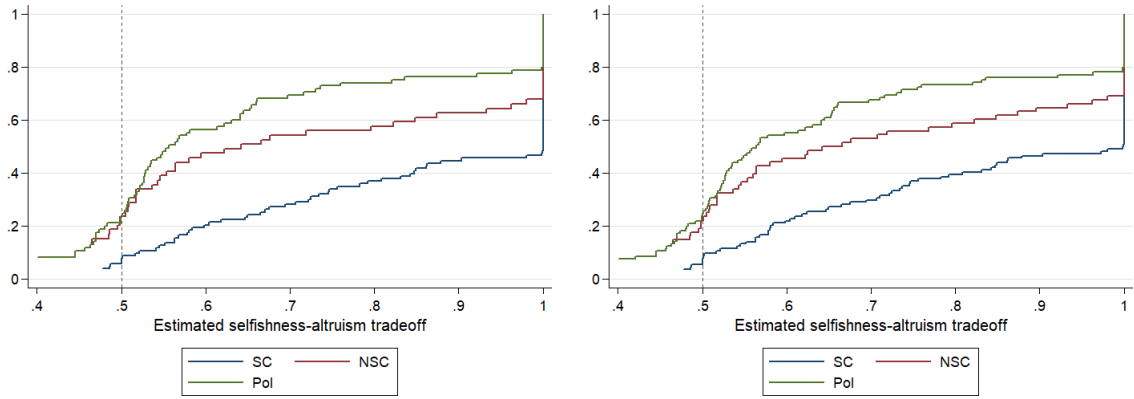
In this section, we plot the Cumulative Distribution Functions of both parameters ($\hat{\alpha}_s$ and $\hat{\rho}_s$) for each restriction strategy and each experimental group.

The figure B.9 (resp. B.10) plots the CDF of $\hat{\alpha}_s$ for the main groups (resp. Politicians' sub-groups) and for all restriction strategies.

The figure B.11 (resp. B.12) plots the CDF of $\hat{\rho}_s$ for the main groups (resp. Politicians' sub-groups) and for all restriction strategies.

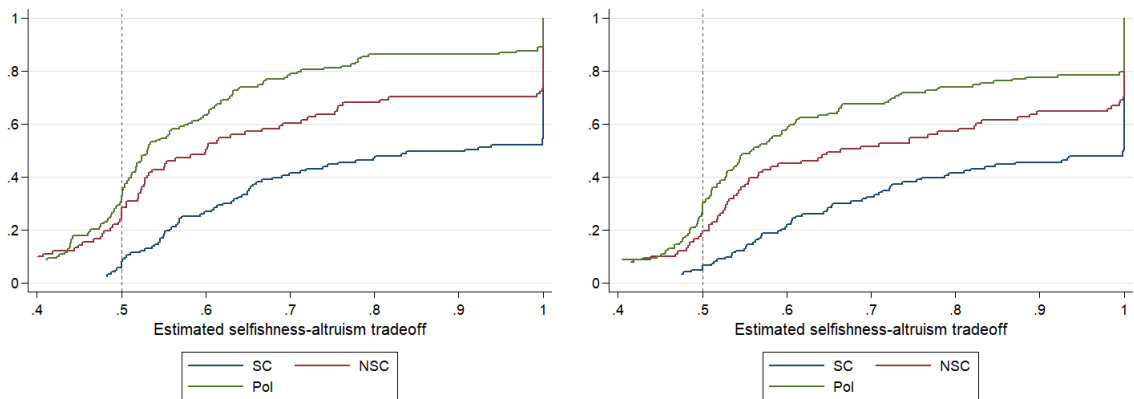
The figure B.13 plots the CDF for both parameters for all groups when no restrictions on homothetic preferences are implemented.

Figure B.9: CDF of estimated Selfishness/Altruism parameter ($\hat{\alpha}_s$) by groups



(a) $HEI \geq 0.51$

(b) Average $HEV \geq 0.927$

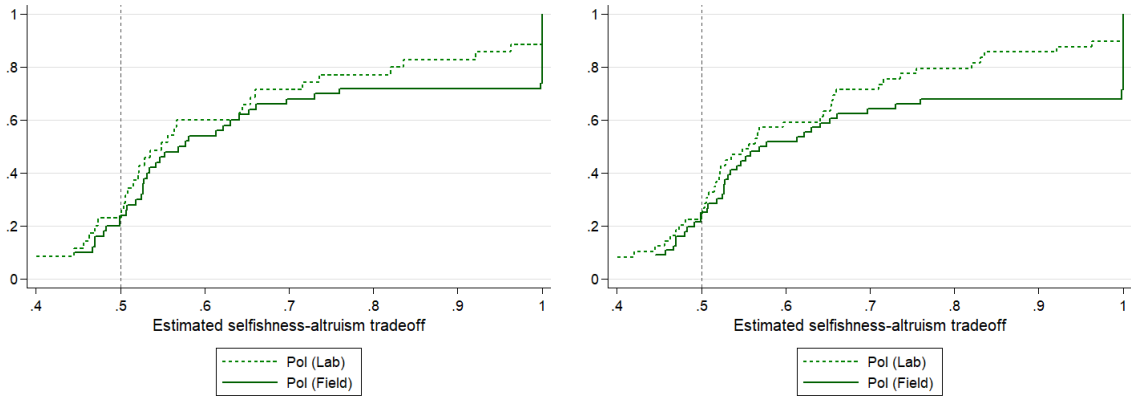


(c) H^2MI

(d) HEV

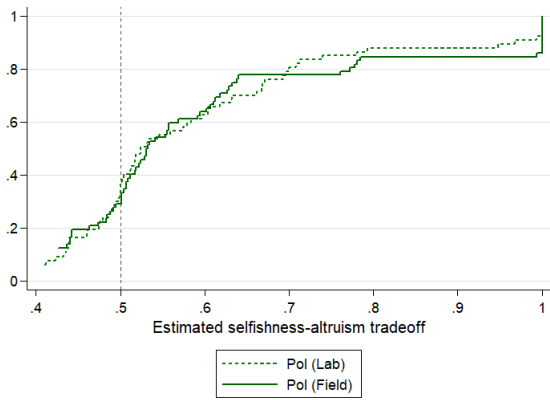
Note: (B.9a) 16 out of 247 subjects have $\hat{\alpha}_s \leq 0.4$: 6 Politicians (2 from the Lab, 4 from the Field), 7 Non-students and 3 Students.
 (B.9b) 18 out of 292 subjects have $\hat{\alpha}_s \leq 0.4$: 7 Politicians (3 from the Lab, 4 from the Field), 8 Non-students and 3 Students.
 (B.9c) 21 out of 353 subjects have $\hat{\alpha}_s \leq 0.4$: 11 Politicians (3 from the Lab, 8 from the Field), 8 Non-students and 2 Students.
 (B.9d) 20 out of 352 subjects have $\hat{\alpha}_s \leq 0.4$: 11 Politicians (4 from the Lab, 7 from the Field), 6 Non-students and 3 Students.

Figure B.10: CDF of estimated Selfishness/Altruism parameter ($\hat{\alpha}_s$) by Politicians' sub-groups

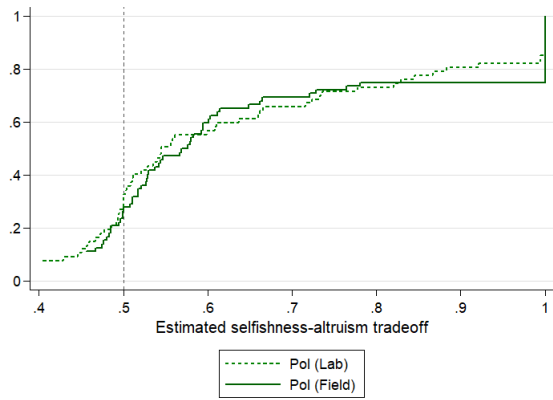


(a) $HEI \geq 0.51$

(b) Average $HEV \geq 0.927$

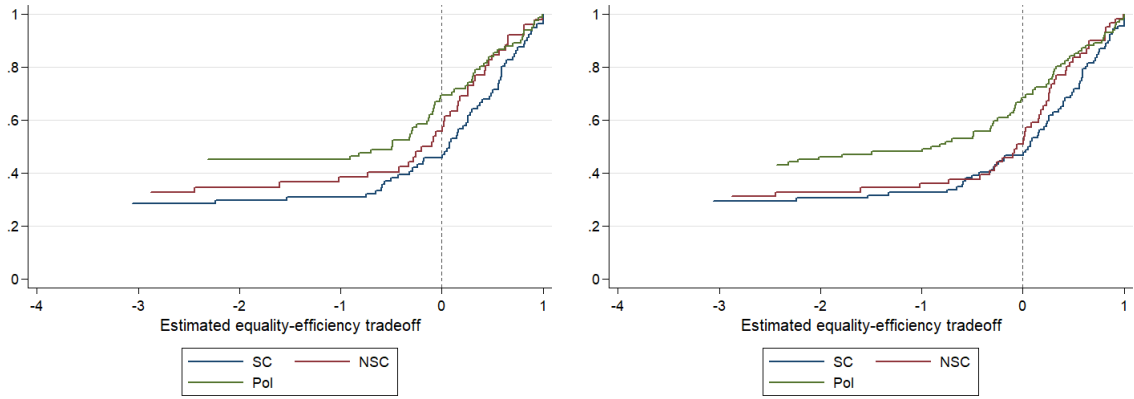


(c) H^2MI



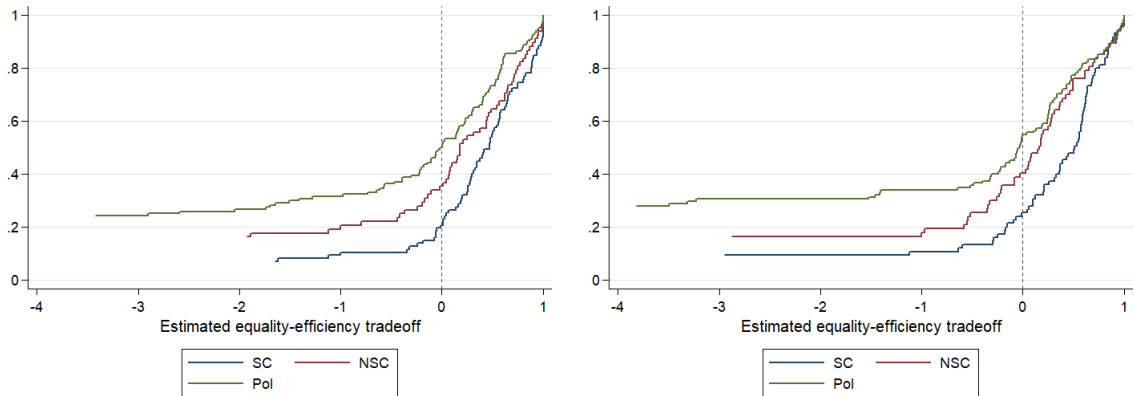
(d) HEV

Figure B.11: CDF of estimated Equality/Efficiency parameter ($\hat{\rho}_s$) by groups



(a) HEI ≥ 0.51

(b) Average HEV ≥ 0.927

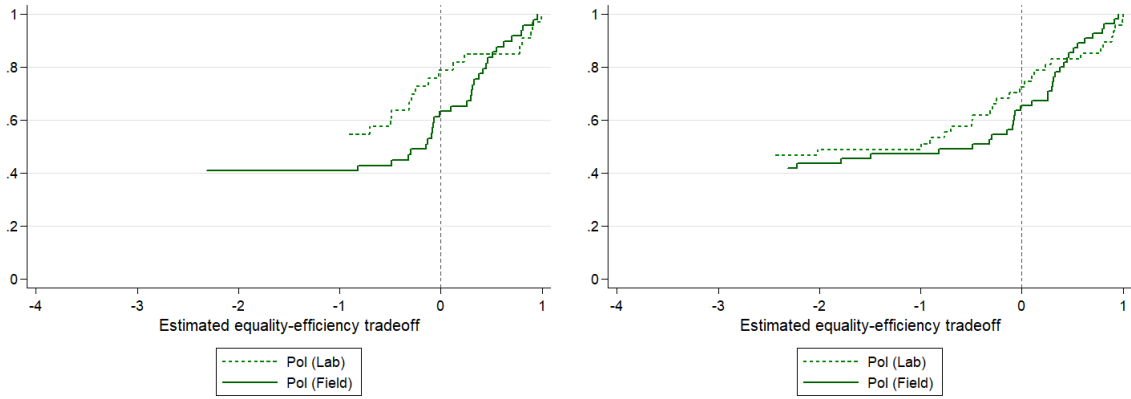


(c) H²MI

(d) HEV

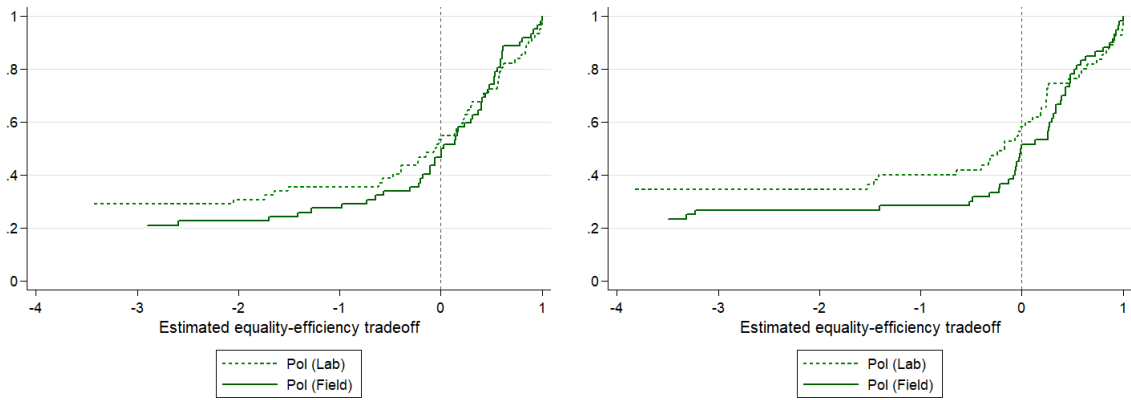
Note: (B.11a) 74 out of 215 subjects have $\hat{\rho}_s \leq -4$: 36 Politicians (17 from the Lab, 19 from the Field), 16 Non-students and 22 Students.
 (B.11b) 87 out of 260 subjects have $\hat{\rho}_s \leq -4$: 43 Politicians (21 from the Lab, 22 from the Field), 18 Non-students and 26 Students.
 (B.11c) 44 out of 279 subjects have $\hat{\rho}_s \leq -4$: 29 Politicians (17 from the Lab, 12 from the Field), 10 Non-students and 5 Students.
 (B.11d) 47 out of 257 subjects have $\hat{\rho}_s \leq -4$: 31 Politicians (18 from the Lab, 13 from the Field), 10 Non-students and 6 Students.

Figure B.12: CDF of estimated Equality/Efficiency parameter ($\hat{\rho}_s$) by Politicians' sub-groups



(a) $HEI \geq 0.51$

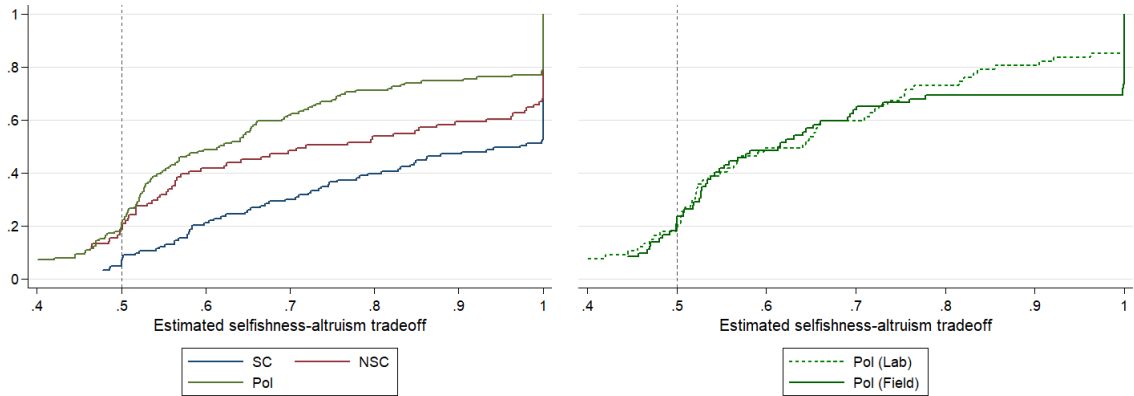
(b) Average $HEV \geq 0.927$



(c) H^2MI

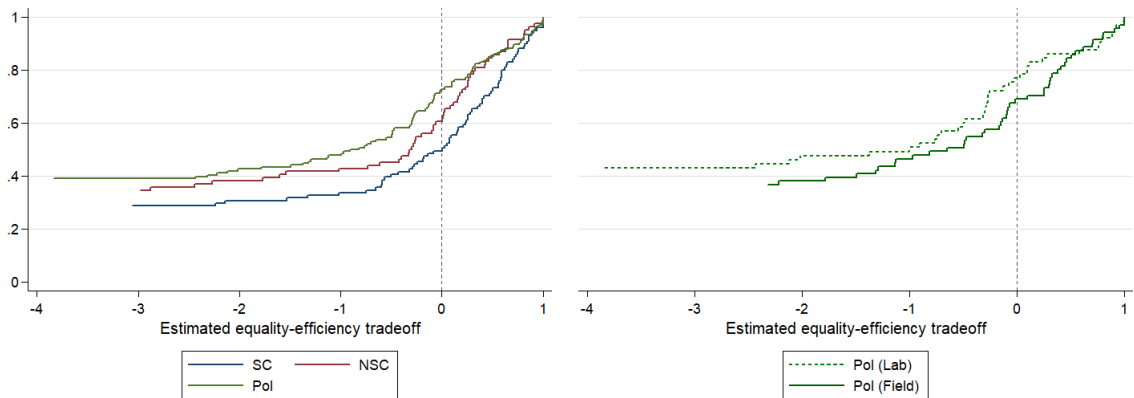
(d) HEV

Figure B.13: CDF of estimated parameters without restrictions on homothecity



(a) Main groups ($\hat{\alpha}_s$)

(b) Politicians' sub-groups ($\hat{\alpha}_s$)



(c) Main groups ($\hat{\rho}_s$)

(d) Politicians' sub-groups ($\hat{\rho}_s$)

Note: 22 out of 353 subjects have $\hat{\alpha}_s \leq 0.4$: 9 Politicians (4 from the Lab, 5 from the Field), 10 Non-students and 3 Students.
 108 out of 353 subjects have $\hat{\rho}_s \leq -4$: 52 Politicians (27 from the Lab, 25 from the Field), 28 Non-students and 28 Students.

The table B.5 exposes p-values from statistical tests comparing the distributions of estimated parameters and the proportions of some preference types as defined by values of the parameters between the experimental groups when no no restrictions on homothetic preferences are imposed.

Table B.5: p -values from statistical tests for differences in estimated parameters (no restrictions)

	$\hat{\alpha}_s$			Prop $\hat{\alpha}_s \geq 0.95$		Prop $\hat{\alpha}_s \in (0.45, 0.55)$		Prop $\hat{\alpha}_s < 0.45$		$\hat{\rho}_s$			Prop $\hat{\rho}_s < 0$	
	t	FP	KS	Pr	F	Pr	F	Pr	F	t	FP	KS	Pr	F
Pol v. NSC ($n = 230/220$)	.2170951	.1103738	.0250632	.0104662	.0124973	.073156	.0959009	.6858187	.822642	.6334411	.4915699	.3578618	.0616364	.0740467
Lab v. Field ($n = 139/136$)	.6426412	.713496	.3646736	.0503012	.071854	.6591495	.7169862	.6687923	.7741648	.1198466	.5485795	.4680111	.300538	.3386718

Note: t: two-samples Student test. FP: Fligner-Policello. KS: Kolmogorov-Smirnov. Pr: Test of proportions. F: Fisher's exact test.

n indicates the number of observations for $\hat{\alpha}_s$ (resp. $\hat{\rho}_s$) on the left (resp. right).

B.4.2 Correlations between estimated parameters

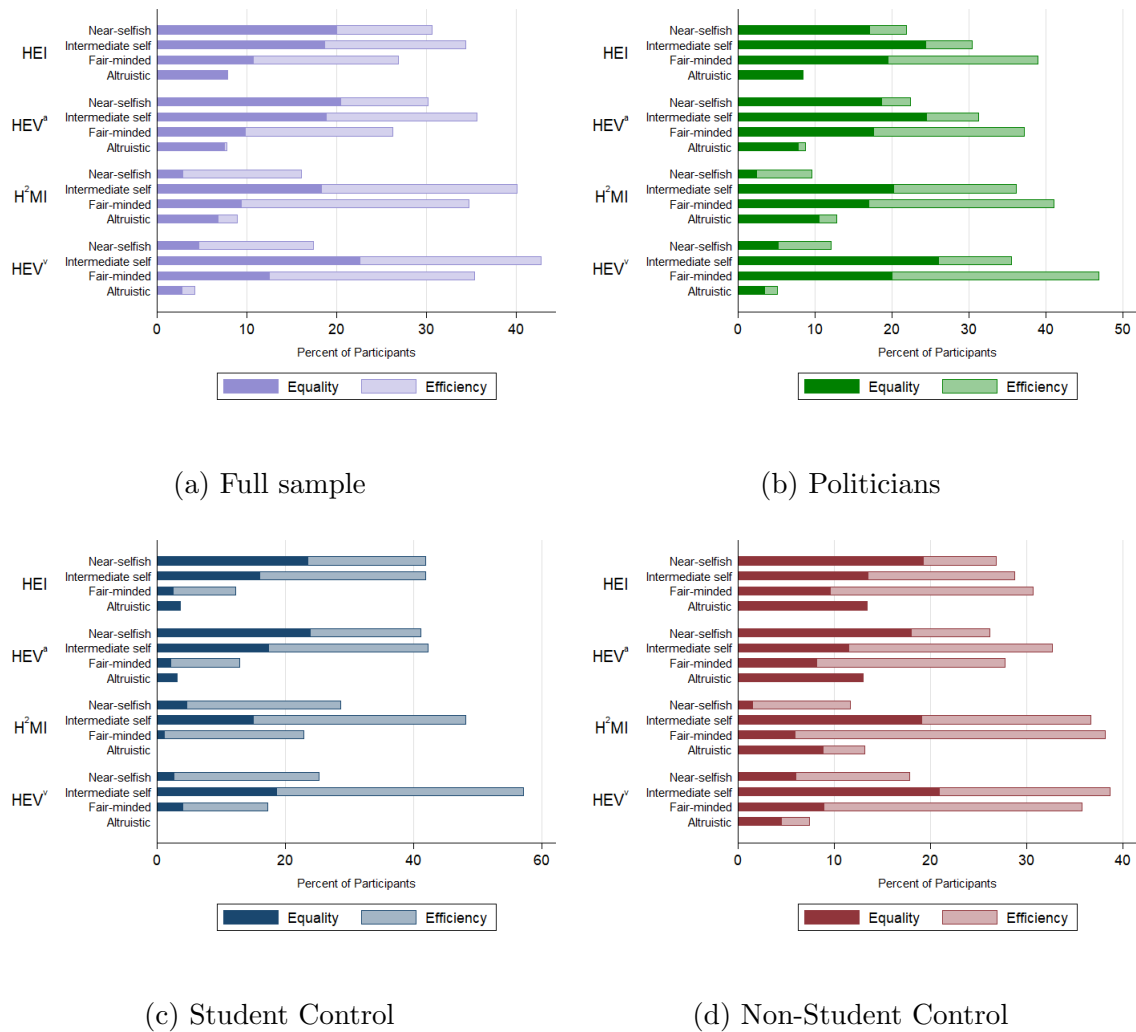
In this section, we look at the within-subject correlation between estimated parameters. Given the importance of outliers in the distribution of parameters (especially $\hat{\rho}_s$), which will likely bias simple correlation coefficients, we focus on a typology of subjects (Fisman et al., 2015b, 2017). The Selfishness/Altruism parameter is divided into 4 categories: near-selfishness ($\hat{\alpha}_s \geq 0.95$), intermediate selfish-orientation ($\hat{\alpha}_s \in (0.55, 0.95)$), fair-mindedness ($\hat{\alpha}_s \in (0.45, 0.55)$) and altruistic-orientation ($\hat{\alpha}_s \leq 0.45$). The Equality/Efficiency parameter is divided into 2 categories: equality-orientation ($\hat{\rho}_s < 0$) and efficiency-orientation ($\hat{\rho}_s > 0$). The figure B.14 summarizes the data for each restriction strategy.

All Fisher’s exact tests strongly reject the independence assumption between the $\hat{\alpha}_s$ and the $\hat{\rho}_s$ categories ($p < 0.01$). Yet, the overall pattern is not clear-cut. First, we observe equality-seekers and efficiency-seekers in all categories of $\hat{\alpha}_s$, with the exception of altruism-orientation in the case of subject-based restrictions. Second, though altruistic subjects appear much more likely to be also classified as equality-oriented, the converse is true for fair-minded subjects. Third, the predominance of efficiency-oriented subjects among selfish types is not obvious: it is even the contrary for subject-based restrictions.¹⁶

As a consequence, we cannot conclude in a straightforward correlation between estimated parameters and still observe substantial heterogeneity in individual profiles.

¹⁶We observe similar results when looking at each experimental group separately.

Figure B.14: Correlation between parameter-based classifications

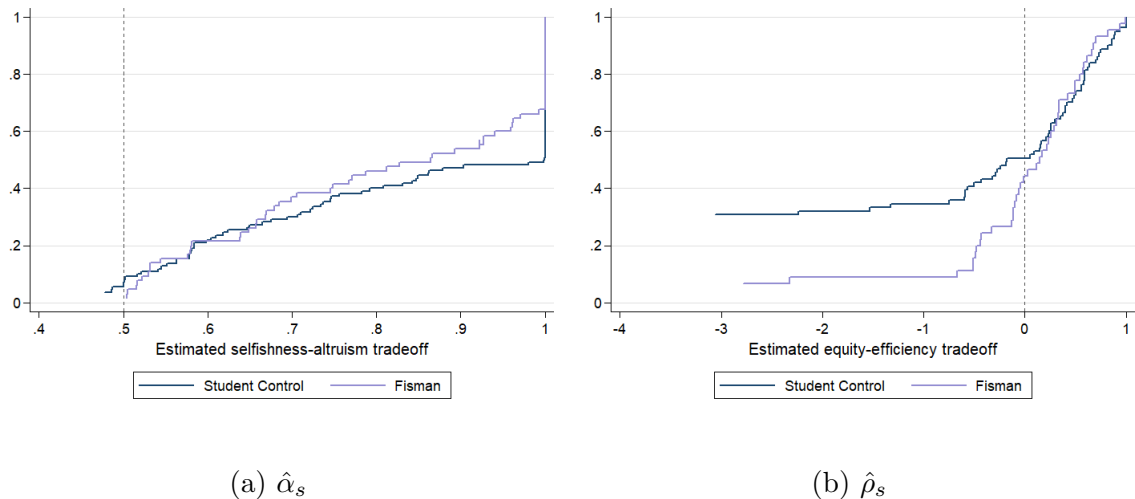


Note: Near-selfishness: $\hat{\alpha}_s \geq 0.95$; Intermediate self-orientation: $\hat{\alpha}_s \in (0.55, 0.95)$; Fair-mindedness: $\hat{\alpha}_s \in (0.45, 0.55)$; Altruistic-orientation: $\hat{\alpha}_s \leq 0.45$; Equality-orientation: $\hat{\rho}_s < 0$; Efficiency-orientation: $\hat{\rho}_s > 0$

B.4.3 Comparisons with Fisman et al. (2007)'s data

In this sub-section, we compare our results on Students with the results reported by Fisman et al. (2007). The figure B.15 plots the cumulative distribution functions of the parameters.¹⁷

Figure B.15: CDF of estimated parameters



Note: (B.15a) 3 out of 175 subjects have $\hat{\alpha}_s \leq 0.4$: 3 Students.

(B.15b) 26 out of 126 subjects have $\hat{\rho}_s \leq -4$: 24 Students and 2 subjects from Fisman et al. (2007).

First, we observe little differences between the two student samples regarding the Selfishness/Altruism tradeoff until $\hat{\alpha}_s \approx 1$, with a bigger proportion of strictly selfish students within our sample (50.91% against 33.85% with average $\hat{\alpha}_s \geq 0.99$). These observations are supported by the statistical analysis. All tests on central tendencies or entire distributions fail to reject the null of no difference (t-test: $p = 0.771$; FP

¹⁷There are two methodological differences between our study and Fisman et al. (2007). First, they exclude all subjects with a Critical Cost Efficiency Index (CCEI) < 0.8 , hence subjects whose behavior sufficiently deviate from coherent preferences (rationality), whereas we go a step further and restrict our attention to subjects whose behavior is sufficiently consistent with homothetic preferences. Second, they use 0.95 as a threshold on the average hold rate to classify subjects as "strictly selfish" and exclude them from the estimation process, whereas we use the 0.99 threshold that is used in earlier studies (e.g., Fisman et al., 2015a, 2017). For the sake of comparability, we neutralize such differences by implementing the 0.95 threshold and by focusing on Students with CCEI ≥ 0.8 .

test: $p = 0.353$; KS test: $p = 0.118$), though the difference in proportion of subjects with average $\hat{\alpha}_s \geq 0.99$ is significant (test of proportions: $p = 0.028$; Fisher's exact test: $p = 0.04$).

Second, we observe that the CDF of $\hat{\rho}_s$ from Students is above the one from the subjects in Fisman et al. (2007)'s study until approximately 0.3. In other words, our student sample appears more equality-oriented than the original student sample. More specifically, there are 49.38% subjects in the former that are efficiency-oriented ($\hat{\rho}_s > 0$) against 55.56% in the latter. We further observe that many more subjects have lower values of the parameter in our experiment, which generates a much longer trail. The difference thus does not concern the number of subjects identified as being equality or efficiency-oriented but rather the intensity of the equality-orientation. Indeed, only the difference in entire distributions is significant (KS test: $p = 0.012$).¹⁸

¹⁸FP test: $p = 0.101$; Test of proportions on $\hat{\rho}_s > 0$: $p = 0.507$; Fisher's exact test on $\hat{\rho}_s > 0$: $p = 0.578$.

B.4.4 Distributions of prediction errors

In this sub-section, we investigate the error distributions from the estimation process. Specifically, we compute the thirty predicted hold rates that we should have observed for each subject given her CES estimated parameters and the relative prices. Then we subtract these predicted values to the actual hold rates in order to obtain the prediction error for each observation. We repeat the process for every restriction strategy. Finally, we plot the distributions of prediction errors (figure B.16, B.17 and B.18) and run usual statistical tests (table B.6). In particular, we are concerned by possible differences in error distributions across groups, which would in turn likely bias the relationships across groups about the estimated parameters.¹⁹

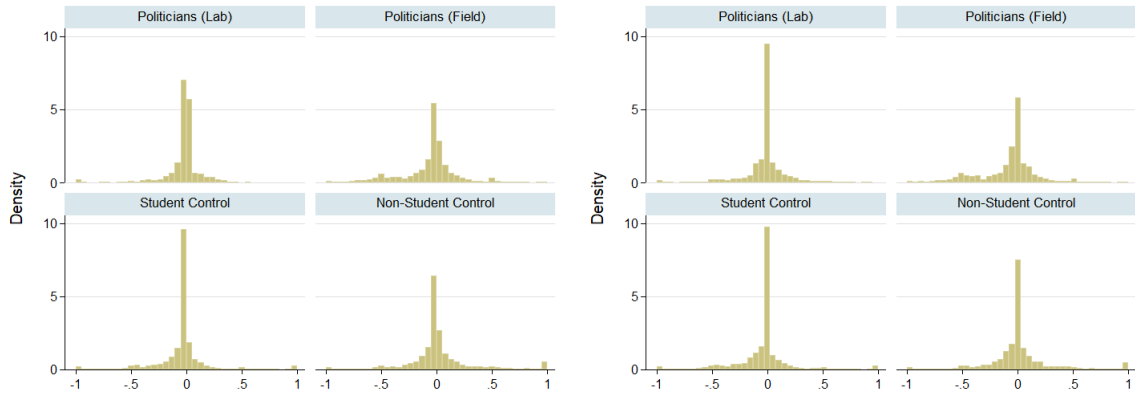
A visual overview based on figures B.16, B.17 and B.18 provides several insights. First, the error distributions are globally consistent for each group across restriction strategies, which means that our strategies are overall performing equally well.²⁰ Moreover, all distributions are normal-shaped with a mode around zero, which we interpret as a sign that our model is actually a good predictor of observed behavior.

Second, the distribution for Students is globally more concentrated than the distribution for both other groups. Specifically, the CDF for Students is generally below (resp. above) the CDF for Politicians and Non-Students until (resp. past) 0. In other words, the prediction errors for Students are lower than for both other groups. Such a result is unsurprising considering that Students are closer to (ho-

¹⁹Specifically, in a regression model of estimated parameters it would mean that the variable identifying the groups and the residuals (which would contain the prediction error) are not independent, hence a violation of the usual zero conditional mean assumption.

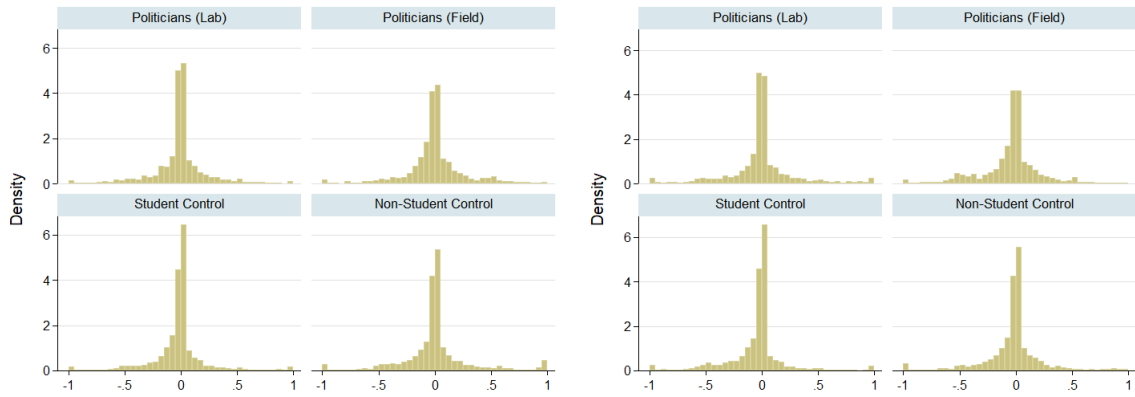
²⁰For observations-based restriction strategies (i.e., based on the H²MI or the HEV), the procedure implies out-of-sample predictions. Indeed, we evaluate the prediction errors for all observations and not only for the observations that are used in the estimation process. Consequently, such a result means that we are not generating many more errors by modelling behavior for all subjects based only on a subset of homothecity-consistent decisions as opposed to modelling behavior only for homothecity-consistent subjects based on all their decisions. This is additional evidence that we can safely rely on observations-based restriction strategies as a complement to subjects-based restriction strategies in order to take into account deviations from the underlying utility model without discarding too many subjects from the analysis.

Figure B.16: Prediction error of hold rate by groups



(a) $HEI \geq 0.51$

(b) Average $HEV \geq 0.927$



(c) H^2MI

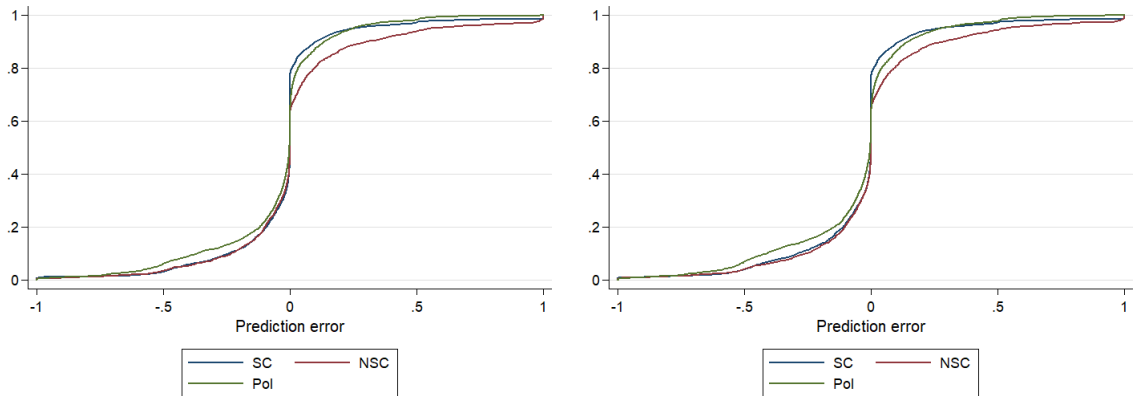
(d) HEV

thetic) utility maximization than both other groups (see the chapter 1 and the sub-section B.1.2 from the appendix B).

Third, though the difference in distributions between Politicians and Non-Students is less conspicuous, the CDF for Politicians tend to be above the CDF for Non-Students. This entails that, relative to Non-Students, our model is under-estimating the pass rate of Politicians, hence the weight they place on others. In other words, the difference in predicted values between Politicians and Non-Students is actually lower than the observed difference, hence a conservative stance with respect to our results.

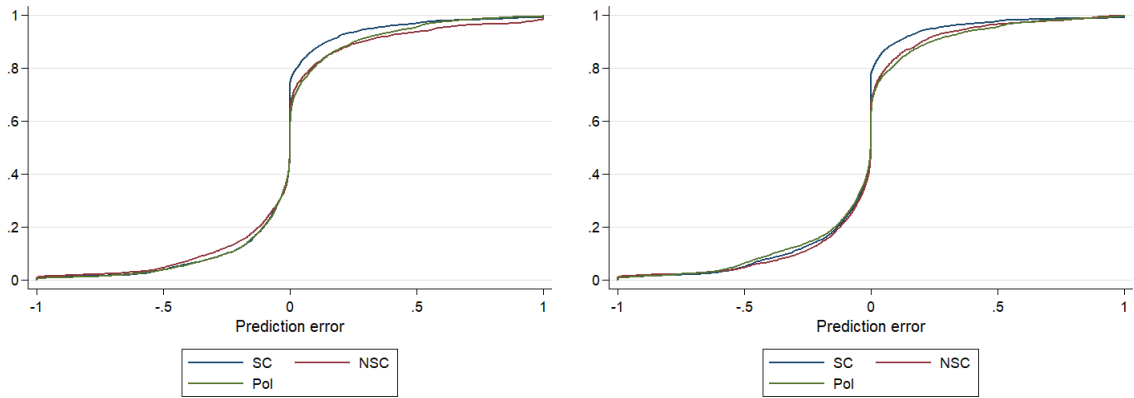
Fourth, the distribution for Politicians from the Lab is more concentrated than

Figure B.17: CDF of prediction error of hold rate by groups



(a) $HEI \geq 0.51$

(b) Average $HEV \geq 0.927$



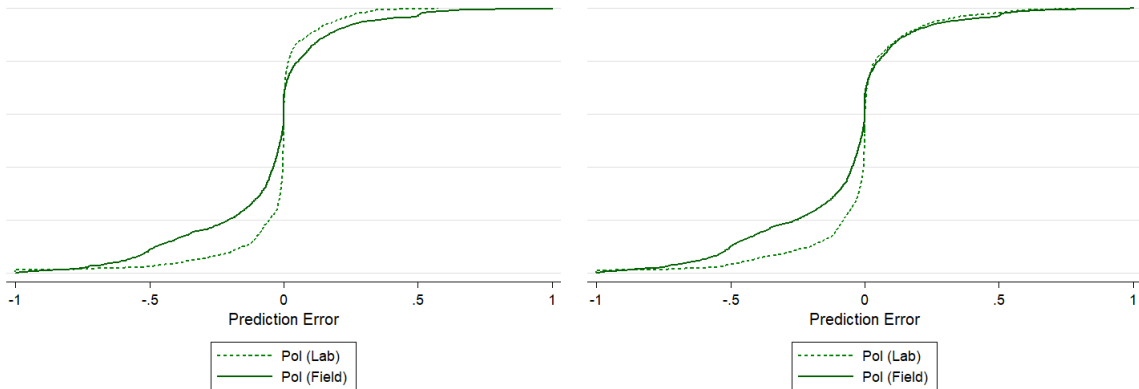
(c) H^2MI

(d) HEV

the distribution for Politicians from the Field. Though the difference is moderated for observations-based restriction strategies, it implies that the model performs better at predicting the behavior of the former group as compared to the latter. It also means that the comparisons between the two Politicians' sub-groups may be confounded by differences in prediction errors, which may thus explain why we observe some differences in the parameters between the two groups.

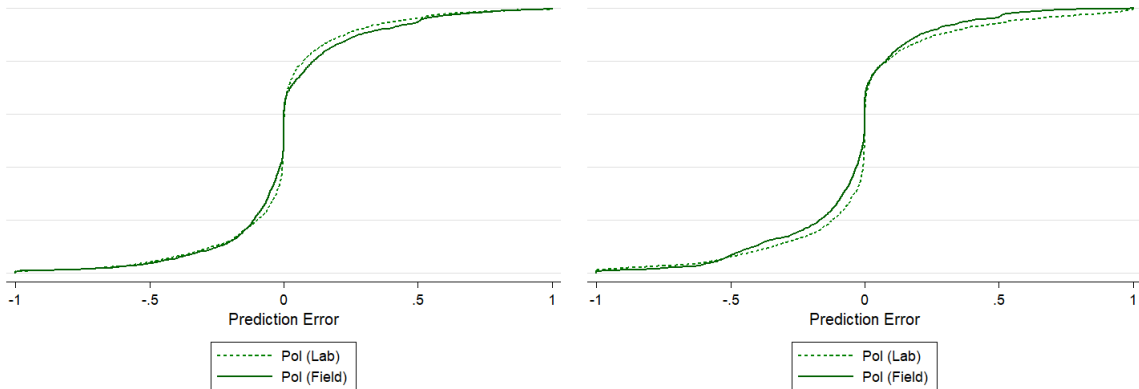
Statistical tests on differences in the distributions of prediction errors between our groups generally reject the null hypothesis that the distributions are the same,

Figure B.18: CDF of prediction error of hold rate by Politicians' sub-groups



(a) $HEI \geq 0.51$

(b) Average $HEV \geq 0.927$



(c) H^2MI

(d) HEV

with the exception of some tests related mostly to observations-based restriction strategies (table B.6).

However, we run these tests on all prediction errors, hence considering a very large number of observations. Consequently, the likelihood to detect even a marginal difference is actually high. Besides, we neglect that prediction errors are likely correlated within subjects. Additionally, there is a comparability issue. Testing differences in prediction errors aims at checking whether differences in parameters might be biased. Yet, the latter differences are based on comparing groups with only one value per subject. We should thus prefer a similar comparison involving only a single value per subject for the prediction errors as well.

Table B.6: Summary statistics and statistical tests on prediction error of hold rate

(a) Summary statistics

	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
HEI	-0.101	-0.021	-0.099	0.069	-0.082	0.032
	(0.184)	(0.372)	(0.187)	(0.208)	(0.213)	(0.307)
	[-0.100]	[-0.306]	[-0.258]	[-0.009]	[-0.165]	[-0.097]
	[-0.005]	[-0.018]	[-0.033]	[0.000]	[-0.006]	[0.000]
	[0.003]	[0.297]	[0.003]	[0.178]	[0.000]	[0.084]
HEV ^a	-0.121	-0.030	-0.082	0.073	-0.105	0.017
	(0.254)	(0.353)	(0.334)	(0.246)	(0.230)	(0.311)
	[-0.349]	[-0.266]	[-0.342]	[-0.014]	[-0.203]	[-0.151]
	[-0.005]	[-0.018]	[-0.041]	[0.005]	[-0.011]	[0.000]
	[0.028]	[0.279]	[0.030]	[0.195]	[0.000]	[0.084]
H ² MI	-0.094	0.044	-0.077	0.086	-0.097	0.019
	(0.196)	(0.354)	(0.377)	(0.351)	(0.262)	(0.349)
	[-0.220]	[-0.226]	[-0.220]	[-0.049]	[-0.192]	[-0.115]
	[-0.014]	[-0.000]	[-0.034]	[0.000]	[-0.007]	[0.000]
	[-0.000]	[0.303]	[0.044]	[0.206]	[0.000]	[0.103]
HEV ^v	-0.048	-0.088	-0.057	0.039	-0.111	0.001
	(0.329)	(0.371)	(0.343)	(0.281)	(0.238)	(0.342)
	[-0.305]	[-0.366]	[-0.216]	[-0.077]	[-0.236]	[-0.151]
	[0.001]	[-0.071]	[-0.017]	[0.000]	[-0.037]	[0.000]
	[0.044]	[0.108]	[0.081]	[0.170]	[0.000]	[0.098]

(b) Statistical tests

	Pol v. NSC				Pol v. SC				Lab v. Field			
	HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v
t-test	<0.01	<0.01	.929	.511	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	.205	<0.01
FP test	<0.01	<0.01	.148	.459	.544	.828	<0.01	<0.01	<0.01	<0.01	.502	<0.01
KS test	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Observations	4316	5186	6895	6895	5632	6563	7851	7851	2546	3146	4165	4165

Note: (B.6a) Entries report mean, standard deviation (parentheses) and quartiles (brackets) of prediction error (all decisions).

(B.6b) Entries report p-value of the corresponding test.

HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

Alternatively, we run the same tests on the average prediction error computed at the individual level and find rather opposite results: not rejecting the null of equal distributions becomes the most common finding (table B.7).²¹

In particular, all non-parametric tests fail to reject the null between Politicians and Non-Students, and only some of the parametric tests detect a significant difference. Such a difference seems to be driven by Politicians from the Field for which the average prediction error is significantly lower on average than the one from Non-Students for both subjects-based restriction strategies (t-test: $p = 0.05$ (HEI) and $p = 0.022$ (HEV^a)). On the contrary, Politicians from the lab and Non-Students do not differ (t-test: $p = 0.295$ (HEI) and $p = 0.43$ (HEV^a)). In light of these observations, we run an additional batch of tests comparing Politicians from the Lab and Non-Students (table B.8). Globally, we uncover statistically stronger results as compared to the ones exposed in tables 2.2 and 2.3: the differences that were significant remain significant and some differences that were not significant are significant (in particular for the restrictions based on the average in the HEV). Together with the fact that there are little statistical differences in estimated parameters between the two Politicians' sub-groups, we interpret such a finding as a confirmation that pooling Politicians is likely to provide at worst conservative estimates. Consequently, it seems unlikely that observed differences in estimated parameters between Politicians and Non-Students are driven by different distributions of prediction errors.

The statistical patterns for comparing Politicians and Students as well as the two Politicians' sub-groups are less clear-cut. In particular, though most parametric tests fail to reject the null of no difference overall, almost two-thirds of the non-parametric tests do reject it. We therefore cannot conclude on the irrelevance of a possible bias on our results on $\hat{\alpha}_s$ and $\hat{\rho}_s$ due to different distributions of the error term between such groups.

²¹Furthermore, the distribution of average prediction errors is more concentrated and more centered around 0 than the distribution of prediction errors for all groups. In other words, the model performs well even within subjects.

However, we note that differences in prediction errors between Politicians and Students were expected given the latter's higher proximity with homothetic preferences. Besides, the important statistical strength of our results on both parameters between these groups in combination with the moderate differences we observe on prediction errors casts some doubts about the latter fully explaining the former.

Finally, the fact that differences in prediction errors may introduce some noise in the comparisons of parameters between Politicians from the Field and Politicians from the Lab actually supply a possible explanation for why we observe moderate differences in parameters between the two groups. It may simply be that we are more able to accurately model the observed behavior in the Lab than in the Field. Ultimately, since Politicians from the Field are closer to Non-Students than Politicians from the Lab in both parameters, the fact that the model is better at predicting observed behavior for the latter ultimately entails a conservative stance. If we do observe significant differences between Politicians and Non-Students in our experiment, it is thus likely that the differences in the general population are ultimately wider.

Table B.7: Summary statistics and statistical tests on average prediction error of hold rate

(a) Summary statistics

	Pol (Lab)	Pol (Field)	Pol (Total)	NSC	SC	Total
HEI	-0.039	-0.030	-0.087	0.019	-0.053	-0.009
	(0.129)	(0.287)	(0.183)	(0.210)	(0.144)	(0.197)
	[-0.005]	[-0.035]	[-0.039]	[-0.004]	[-0.066]	[-0.020]
	[-0.000]	[-0.000]	[-0.000]	[-0.001]	[-0.006]	[-0.002]
	[0.000]	[0.001]	[0.000]	[0.000]	[-0.000]	[0.000]
HEV ^a	-0.039	-0.053	-0.075	0.005	-0.056	-0.017
	(0.216)	(0.290)	(0.171)	(0.213)	(0.145)	(0.197)
	[-0.008]	[-0.046]	[-0.022]	[-0.004]	[-0.066]	[-0.025]
	[-0.000]	[-0.001]	[-0.000]	[-0.001]	[-0.006]	[-0.002]
	[0.001]	[0.002]	[0.000]	[0.000]	[-0.000]	[0.001]
H ² MI	-0.055	0.052	-0.036	0.004	-0.041	0.003
	(0.148)	(0.204)	(0.157)	(0.211)	(0.140)	(0.157)
	[-0.114]	[-0.046]	[-0.057]	[-0.031]	[-0.086]	[-0.047]
	[-0.002]	[0.000]	[-0.000]	[-0.000]	[-0.011]	[-0.000]
	[0.007]	[0.034]	[0.026]	[0.034]	[-0.000]	[0.022]
HEV ^v	0.010	-0.067	-0.032	-0.031	-0.067	-0.027
	(0.255)	(0.275)	(0.189)	(0.113)	(0.151)	(0.169)
	[-0.012]	[-0.054]	[-0.028]	[-0.030]	[-0.099]	[-0.053]
	[-0.000]	[-0.014]	[-0.000]	[-0.002]	[-0.013]	[-0.001]
	[0.026]	[0.009]	[0.014]	[0.003]	[-0.000]	[0.010]

(b) Statistical tests

	Pol v. NSC				Pol v. SC				Lab v. Field			
	HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v
t-test	.047	.046	.977	.853	.344	.373	.233	.355	.282	.071	.72	.268
FP test	.508	.68	.852	.377	.16	.209	.047	.012	.049	<0.01	.259	.016
KS test	.622	.838	.796	.455	.125	.112	.02	<0.01	.069	.028	.35	<0.01
Observations	144	173	230	230	188	219	262	262	85	105	139	139

Note: (B.7a) Entries report mean, standard deviation (parentheses) and quartiles (brackets) of prediction error averaged at the individual level.

(B.6b) Entries report p-value of the corresponding test.

HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

Table B.8: Statistical tests for differences in estimated parameters without Field data

(a) Selfishness/Altruism tradeoff $\hat{\alpha}_s$

		Pol v. NSC				Pol v. SC			
		HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v
$\hat{\alpha}_s$	t-test	.324	.194	.048	.055	<0.01	<0.01	<0.01	<0.01
	FP test	.114	.065	.028	.024	<0.01	<0.01	<0.01	<0.01
	KS test	.175	.05	.032	.127	<0.01	<0.01	<0.01	<0.01
Prop $\hat{\alpha}_s \geq 0.95$	Test of proportions	.026	<0.01	<0.01	.017	<0.01	<0.01	<0.01	<0.01
	F-test	.032	<0.01	<0.01	.02	<0.01	<0.01	<0.01	<0.01
Prop $\hat{\alpha}_s \in (0.45, 0.55)$	Test of proportions	.195	.171	.176	.064	<0.01	<0.01	<0.01	<0.01
	F-test	.253	.22	.23	.085	<0.01	<0.01	<0.01	<0.01
Prop $\hat{\alpha}_s < 0.45$	Test of proportions	.949	.937	.712	.909	.047	.014	<0.01	.018
	F-test	1.000	1.000	.823	1.000	.069	.022	<0.01	.035
Observations		94	117	158	158	138	163	190	190

(b) Equality/Efficiency tradeoff $\hat{\rho}_s$

		Pol v. NSC				Pol v. SC			
		HEI	HEV ^a	H ² MI	HEV ^v	HEI	HEV ^a	H ² MI	HEV ^v
$\hat{\rho}_s$	t-test	.019	.048	.984	.471	.017	.021	<0.01	<0.01
	FP test	.398	.362	.079	.074	.078	.121	<0.01	<0.01
	KS test	.137	.073	.168	.066	.012	.034	<0.01	<0.01
Prop $\hat{\rho}_s < 0$	Test of proportions	.03	.024	.04	.049	<0.01	<0.01	<0.01	<0.01
	F-test	.037	.03	.052	.068	<0.01	<0.01	<0.01	<0.01
Observations		85	108	130	122	114	139	149	190

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$. t-test: two-samples Student test. FP: Fligner-Policello. KS: Kolmogorov-Smirnov. F-test: Fisher's exact test.

B.5 Complementary regression analyses

B.5.1 Specification curves for the regressions on the estimated parameters

In this section we present results on our binary group variable of interest from several econometric models ("specifications") of our estimated parameters which vary the number and set of control variables, the restriction strategy (homothetic indices), the threshold on the exclusion restriction (average $p_{s,d}^i \pi_{s,d}^i$ or average $p_{s,r}^i \pi_{s,r}^i \geq 0.99$ or ≥ 0.95) or the underlying estimation technique (Maximum Likelihood or Non-Linear Least Squares).

These results are displayed graphically in figures B.19 and B.21. Each figure is composed of three sub-figures, one for each comparison of interest: Politicians versus Non-Students, Politicians versus Students and Politicians from the Lab versus Politicians from the Field. In each sub-figure, the top part displays the regression coefficient associated with the binary group variable along with 90% and 95% confidence intervals, the middle part displays the corresponding p-value when implementing the score cluster bootstrap correction (Kline and Santos, 2012), and the bottom part identifies the model's specification.

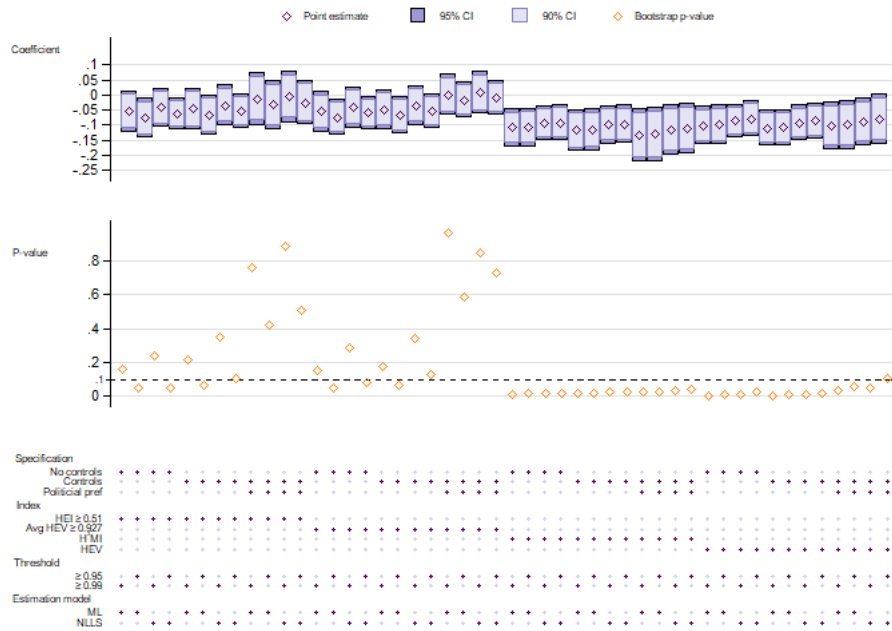
The interpretation of the graph is thus akin to specification curve analysis (Simonsohn et al., 2020): it allows to identify the variations on the coefficient of interest (the group variable) and its statistical significance due to variations in the "specification" of the econometric model. Specifically, when one bar overlaps 0, the corresponding coefficient is not statistically significant at the given confidence level (95% of 90%). The main difference with standard specification curve analysis is that we do not only consider variations in the number and type of included variables in the model. We also consider variations in the way our dependant variables (the individual-level parameters from the structural estimation) are constructed. In addition, we plot a specific graph for the corrected p-value which cannot be inferred from the top graph.

The estimates from comparing Politicians to Students are remarkably close for both parameters, and corresponding p-values are always very low. Similarly, the comparisons involving Politicians' sub-groups yield approximately the same results

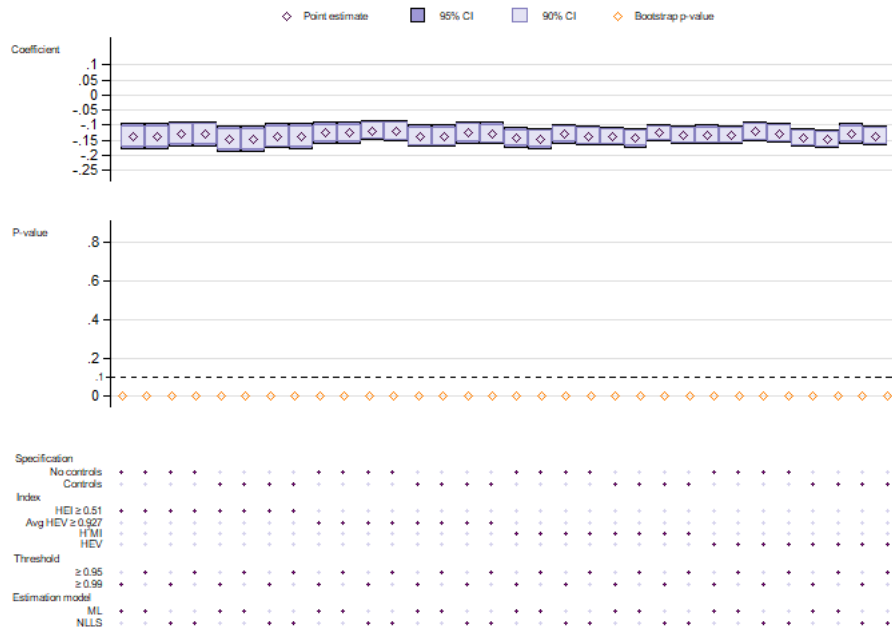
no matter the specification for both parameters. There is one exception though: estimated $\hat{\alpha}_s$ under average HEV restriction often yields a significant difference between the groups, but in most cases the p-value is close to the 10% threshold.

For the Politicians versus Non-Students comparisons, there is more variation in the estimates and p-values, but all coincide with what was previously observed. Specifically, all models relying on observations-based restrictions involve significant results for both parameters. On the contrary, some models involving subjects-based restrictions yield non-significant differences. In particular, it is the case for all models which incorporates political preferences. For $\hat{\rho}_s$, it is also the case when only the binary group variable is introduced with the HEI restriction, but adding control variables reduces the p-value. For $\hat{\alpha}_s$, interestingly, enlarging the exclusion restriction to the 0.95 threshold results in significant effects, whereas effects are not significant with the 0.99 threshold.

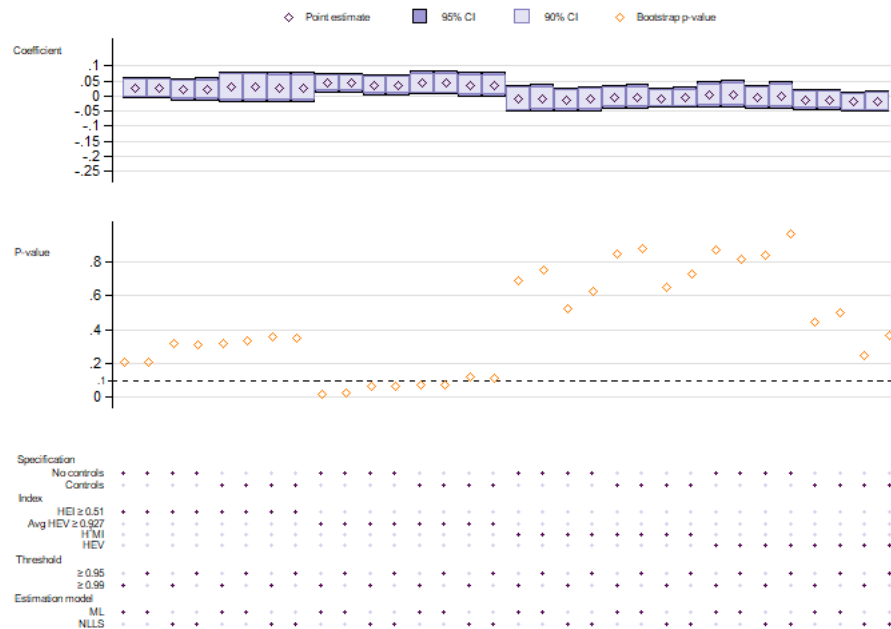
Figure B.19: Specification curves for tobit on estimated Selfishness/Altruism trade-off ($\hat{\alpha}_s$)



Pol v. NSC



Pol v. SC

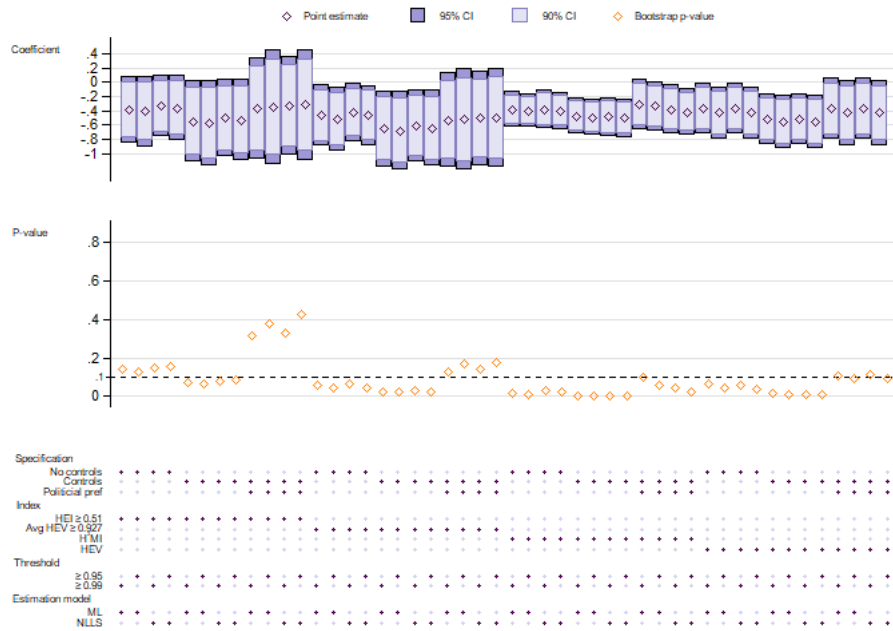


Politicians' sub-groups

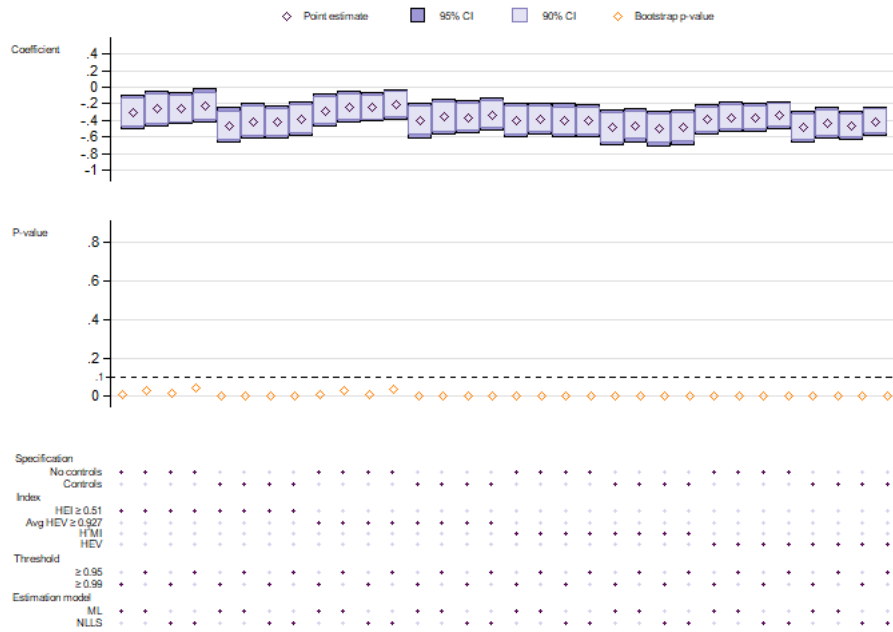
Note: Each graph reports the results from several tobit regressions on $\hat{\alpha}_s$, in the spirit of specification curve analysis (Simonsohn et al., 2020). Specifically, the graph displays the coefficient and corresponding statistical significance associated with the group variable (dummy) which compares either Politicians and a control group or the Politicians' sub-groups. The regressions vary with respect to the inclusion of control variables, the restriction strategy, the exclusion restriction and the structural estimation technique.

Each graph is composed of three parts. The top part plots the regression coefficient along with the 95% and 90% confidence interval. The middle part plots the p-values when implementing the score cluster bootstrap correction (Kline and Santos, 2012). The bottom part identifies the "specification". For each possible specification, a purple bullet identifies whether it is active or not. For instance, the model for the first column includes only the group variable (no control), for subjects with $HEI \geq 0.51$ and relies on $\hat{\alpha}_s$ obtained from a structural model estimated using maximum likelihood and which excludes subjects with average endowment share kept/given above 0.99.

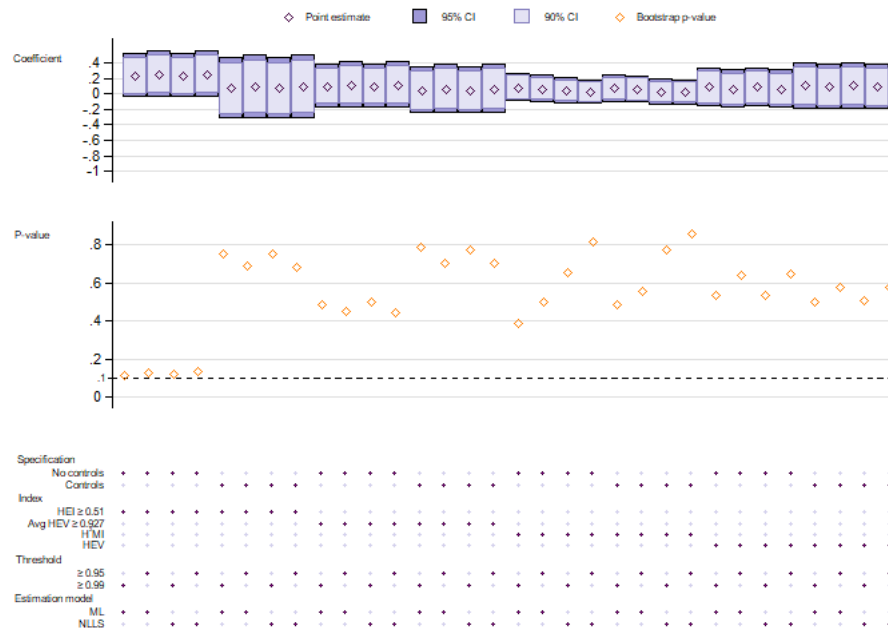
Figure B.21: Specification curves for probit on estimated Equality/Efficiency trade-off ($\hat{\rho}_s > 0$)



Pol v. NSC



Pol v. SC



Politicians' sub-groups

Note: Each graph reports the results from several probit regressions on $\hat{\rho}_s > 0$, in the spirit of specification curve analysis (Simonsohn et al., 2020). Specifically, the graph displays the coefficient and corresponding statistical significance associated with the group variable (dummy) which compares either Politicians and a control group or the Politicians' sub-groups. The regressions vary with respect to the inclusion of control variables, the restriction strategy, the exclusion restriction and the structural estimation technique. Each graph is composed of three parts. The top part plots the regression coefficient along with the 95% and 90% confidence interval. The middle part plots the p-values when implementing the score cluster bootstrap correction (Kline and Santos, 2012). The bottom part identifies the "specification". For each possible specification, a purple bullet identifies whether it is active or not. For instance, the model for the first column includes only the group variable (no control), for subjects with $HEI \geq 0.51$ and relies on $\hat{\rho}_s$ obtained from a structural model estimated using maximum likelihood and which excludes subjects with average endowment share kept/given above 0.99.

B.5.2 Regressions without political preferences in Field v. Non-Students

In this section, we expose the results from regressions using data on Politicians and Non-Students but without including political preferences.

Table B.9: Politicians from the Field v. Non-Students

	HEI		HEV ^a		H ² MI		HEV ^v	
	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$
<u>Pol</u> (Baseline: Control)	-0.007 (0.044)	-0.401 (0.339)	0.009 (0.036)	-0.568* (0.336)	-0.132*** (0.046)	-0.364** (0.183)	-0.103*** (0.040)	-0.380* (0.214)
		[-0.155]		[-0.221]		[-0.141]		[-0.151]
<u>Age</u>	-0.004 (0.006)	-0.024 (0.019)	-0.005 (0.006)	-0.035* (0.018)	-0.004 (0.003)	-0.006 (0.013)	-0.002 (0.003)	-0.004 (0.013)
		[-0.009]		[-0.014]		[-0.002]		[-0.002]
<u>Gender</u> (Baseline: Male)	0.087 (0.057)	-0.539** (0.273)	0.055 (0.047)	-0.477* (0.282)	-0.016 (0.053)	-0.149 (0.270)	0.032 (0.053)	-0.194 (0.239)
		[-0.209]		[-0.186]		[-0.058]		[-0.077]
<u>Higher education</u> (Baseline: No)	0.011 (0.072)	0.255 (0.272)	0.029 (0.078)	0.122 (0.303)	-0.036 (0.061)	0.295 (0.209)	-0.058 (0.047)	0.593** (0.264)
		[0.099]		[0.048]		[0.114]		[0.235]
<u>Income</u> (Baseline: Between 2000 € and 3000 €)								
< 2000 €	-0.140 (0.096)	0.426 (0.331)	-0.120 (0.077)	0.390 (0.291)	-0.004 (0.054)	0.381* (0.224)	-0.081 (0.051)	0.413 (0.255)
		[0.166]		[0.150]		[0.148]		[0.163]
> 3000 €	-0.063 (0.124)	0.098 (0.375)	-0.090 (0.113)	0.361 (0.361)	0.055 (0.087)	0.268 (0.258)	0.037 (0.101)	0.330 (0.248)
		[0.037]		[0.139]		[0.106]		[0.131]
Do not want to answer	-0.099 (0.094)	0.123 (0.468)	-0.117 (0.101)	0.618 (0.512)	-0.038 (0.060)	0.685 (0.439)	-0.090 (0.082)	0.339 (0.552)
		[0.046]		[0.241]		[0.252]		[0.134]
Observations	108	100	124	116	162	130	162	127
LR chi2	7.116	9.229	6.195	10.575	20.900	18.817	42.845	17.441
Prob > χ^2	0.417	0.237	0.517	0.158	0.004	0.009	0.000	0.015
Bootstrapped p-value	0.877	0.236	0.814	0.097	0.030	0.078	0.031	0.106

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses. Marginal effects estimated at the means of covariates in brackets.

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

The "Bootstrapped p-value" report the p-value associated with the group variable when applying the score cluster bootstrap correction for the small number of clusters (Kline and Santos, 2012). The bootstrap is conducted with 4999 repetitions.

B.5.3 Secondary results on control variables

In this section, we discuss the results we obtain on our control variables in the second column for every restriction strategy in the tables 2.4, 2.5 and 2.6. Indeed, that we obtain similar effects as existing findings on other variables is also a pledge of quality in favor of our data.

First, we expect an effect of gender. Indeed, females have been repeatedly found to act more generously in dictator games than their male counterparts (Engel, 2011) and to be more inequality averse (e.g., the review by Croson and Gneezy, 2009). Based on a similar design as ours, both Fisman et al. (2017) and Müller (2019) find that females are not significantly more altruistic than men but that they are significantly more equality-oriented. We globally find consistent results. On the one hand, we never find females to be significantly more altruistic than males. On the other hand, we systematically observe a negative effect of the gender variable on the probability to be efficiency-oriented. Such an effect is always significant in the Politicians v. Students comparisons, whereas the significance varies depending on the implemented restriction strategy in the Politicians v. Non-Students comparisons.²²

Second we expect an effect of age. Giving in dictator games has been found to increase with age (Engel, 2011). There is also some evidence that older adults tend to target an equal division of endowment more often than younger adults (see for instance Bellemare et al., 2008, 2011, Fisman et al., 2017, Müller, 2019, Kerschbamer and Müller, 2020). Our results are somewhat consistent with such empirical evidence: age is negatively correlated with both selfishness and efficiency-orientation in most regressions. However, the effect is rarely significant at conventional levels. Yet, we note that such a non-significant effect of age may stem from a lack of variation in the age variable. Indeed, the dispersion of age in our experiment is lower than in large survey experiments. This is because we cannot include our student

²²One should bear in mind that, considering that the literature usually implement only subjects-based restriction strategies, only the results from the HEI and the HEV^a are readily comparable. Interestingly, these are the restriction strategies for which the results on gender are the most significant.

sample in our regressions with age due to a strict colinearity issue between our group variable and age (all Students are aged below 32).

Alternatively, the observed age effect in the literature on experimental games may also partly be an artifact caused by confounds in the experimental procedures. Indeed, the study by Kettner and Waichman (2016) emphasizes that inexperienced students actually behave in a similar way as elderly subjects, while experienced students act more selfishly. Given that a bit more than a third of our student sample are inexperienced (self-declared) subjects, we try reproducing their result with our data in a dedicated appendix (B.5.4 below). We do find significant positive effects of experience in participating in economic experiments on the Selfishness/Altruism parameter. Interestingly, we also find a positive and very significant effect of having participated once in the past on the probability to be an efficiency-seeker. These effects are specification-sensitive and do not necessarily persist over increased participations, so that further research would be needed to investigate them. Nevertheless, they could have important methodological implications for past and future experimental studies.

Third, we expect an impact of major. Indeed, one classical debate in experimental economics is the overwhelming use of students who happen to be of a very special kind: the one who study economics. Several studies have indeed emphasized that an economic background is likely to influence decisions in general, and in particular in distributional situations like ours. It has been found to increase both selfishness and efficiency consideration (among others: Frank et al., 1993, Fehr et al., 2006, Cappelen et al., 2015, Frigau et al., 2019).²³ Consequently, we expect to find a positive impact of holding an economic major on both parameters. However, we do not find consistent results across models: having a background in what we dubbed "Economics" does not conspicuously nor systematically lead to higher levels of selfishness or more pronounced efficiency considerations as compared to the baseline

²³Quite similar to the possible channels through which Politicians may differ from the general population, there might be a selection effect (students who self-select into economics are more likely to be selfish and efficiency-oriented) or an indoctrination effect (students who endeavour to understand economics courses value more selfishness and efficiency because they precisely correspond to the core assumptions of the economic models). On these interpretations, see Bauman and Rose (2011)

category (Fundamental sciences). In some cases, the sign of the coefficient even lies in the opposite direction and is significant.

Fourth, there also exists some evidence that education may relate to both our tradeoffs. In the representative sample from Netherlands used by Bellemare et al. (2008, 2011), more educated people are more likely to make selfish and inequality-increasing choices while the converse is observed in Germany by Kerschbamer and Müller (2020). Jakiela et al. (2015) finds that higher educational achievement correlates with an increase in 50-50 splits in a dictator game in Kenya. Similarly, Almås et al. (2020) report that higher education is positively linked to equality considerations in the United States, but the authors also underline that such a relationship is not observed in Norway.²⁴ On the one hand, we do not find a clear relationship between our education level variable and the Selfishness/Altruism tradeoff: the sign of the coefficients is not consistent across models and the relationship is never significant. On the other hand, we observe a positive correlation between having a higher educational degree and efficiency considerations. Such a correlation is even significant in some cases.

Fifth, the effect of income may also deserve some comments. There is indeed a large on-going controversy on the moral behavior of economic elites (especially in the aftermath of the 2008 financial crisis). So far the existing evidence is rather mixed. For instance Piff et al. (2010, 2012) document that high social class individuals are less likely to be generous, helpful or trusting and more likely to lie, be greedy or cheat than their low social class counterparts. Trautmann et al. (2013) decompose the various factors composing "social class" (including wealth and education for instance) and find mixed effect of wealth: positive correlation with volunteering and trusting but also positive correlation with cheating in several domains (tax and

²⁴One possible confounding factor could be that education is probably not best captured through a continuous lens: its quality may also matter. For instance, looking at various early childhood interventions on education, Cappelen et al. (2020) find that different education programs affect differently the emergence of distributional preferences. In particular, targeting the child (through a 9-months preschool period) increases the sensitivity to egalitarian outcomes whereas targeting the parents (through a 9-months parenting program) increases the sensitivity of efficient outcomes as measured in economic games several years later. The type of education, which is likely to vary across societies and cultures, may thus matter more to the emergence of specific types of distributional preferences than the number of years of education.

sexual life). In their study on American elites using a similar experimental design and adopting partly an income-based definition of "elites", Fisman et al. (2015b) find elites to be more selfish and more efficiency-oriented than the remainder of the population. An interesting study by Andreoni et al. (2017) uses the misdirected letter technique in a natural field experiment and underlines that rich people do behave more pro-socially than poor people but that such an effect vanishes once taking into account the pressure associated with low income and the differences in the marginal utility of money. The authors indeed conclude that rich people give more simply because they have more income to give.²⁵ There also exists evidence that high-income individuals are more likely to accept inequalities: individuals from high-income regions are less likely to favor the needy recipients than individuals from low-income regions in Cappelen et al. (2013) and richer subjects are more likely to accept inequalities than poorer subjects in Kerschbamer and Müller (2020). However, Fisman et al. (2017) report an opposite result with low-income individuals favoring efficiency more often than middle-income or high-income individuals, while no effect is found on trading off selfishness and altruism.

Regarding the Selfishness/Altruism tradeoff, we observe little difference between our three income classes: most effects are not significant, though low-income subjects appears globally less selfish than middle/high income subjects. Concerning the Equality/Efficiency tradeoff, we find more robust evidence in favor of low income subjects being more efficiency-oriented than the baseline middle-income group, consistent with Fisman et al. (2017).

Lastly, we expect an impact of political preferences. Indeed, there has been several recent studies based on large samples that document a link between distributional and political preferences. Fisman et al. (2017) find that being equality-oriented increases the probability of a vote in favor of Obama in 2012 and of support for the Democratic party in the United States. Almås et al. (2020) find that conservatives in both Norway and the United States are more likely to accept inequalities in spectator conditions. Kerschbamer and Müller (2020) run a large experiment using the German Internet Panel and find that selfish subjects are more likely to

²⁵The interpretation of such a result for a laboratory experiment is however not clear: would high income individual behave more altruistically because they already have plenty at home or would there be no impact whatsoever given that both income types are given the same endowment in the laboratory?

self-report being right-leaning, more likely to vote for a right-wing political party and less willing to favor redistribution. On the contrary, all other distributional types (inequality-averse, altruistic and maximin) all lie at the other extreme of the political spectrum. Earlier studies based on student samples also pointed out such a relationship. Esarey et al. (2012) highlight that conservatives are more likely than liberals to favor their self-interest in a taxation/redistribution game in a laboratory experiment in the United States. Cappelen et al. (2017) document a positive relationship between left-leaning preferences and giving in a laboratory dictator game in Norway. Müller (2019) underlines a relationship between the Selfishness/Altruism parameter and political preferences (right-wing subjects are more likely to be selfish than left-wing subjects) in Germany. However, he does not find a significant relationship with the Equality/Efficiency parameter.

Our analysis of political preferences combines regressions involving the self-placement variable (third column for each restriction strategy in table 2.4) and regressions involving alternative formulations of political preferences (tables B.12 and B.13 and in appendix B.5.5). Regarding the Selfishness/Altruism parameter ($\hat{\alpha}_s$), we do not find any clear pattern. In particular, the results depend upon both the definition of political preferences that we consider and the restriction strategy. We therefore cannot conclude in the existence of any particular relationship between political preferences and the weight placed on others. Regarding the Equality/Efficiency parameter ($\hat{\rho}_s$), we find a positive association between left-leaning preferences and equality-orientation. This result globally stands no matter the definition of political preferences, the implemented restriction strategy or the inclusion of control variables.²⁶ A more detailed discussion on the definition of political preferences we adopt and its impact on the results is proposed in the dedicated appendix B.5.5.

²⁶Yet, the association is not systematically significant (although it is for 42 out of 48 models). In particular, we observe that when subjects that identify themselves with the "Green party" are classified as "left-wing" instead of "center", the relationship gets stronger. We note that classifying Greens as left-wing is rather relevant in the French political landscape. However, further investigation on bigger datasets would be needed to assess the empirical validity of such finding.

B.5.4 Effect of past experience in experiments

The regressions exposed in table B.10 explore the impact of Students’ past experience in experiments on both our estimated parameters. Our ”past experience” variable is directly taken from a question that is standardly added in the end-of-experiment questionnaire. Subjects could answer either that participating in our experiment was their first participation, second participation, or more. For each parameter and each restriction strategy, we display two models: the first model includes only the past experience variable and the second model adds available control variables.²⁷

Table B.10: Effect of past experience in economic experiments on estimated parameters

	HEI				HEV ^a				H ² MI				HEV ^v			
	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$
<u>Past experience</u> (Baseline: Never participated)																
Participated once	0.128** (0.065)	0.098 (0.075)	0.692*** (0.248)	0.674** (0.317)	0.127* (0.068)	0.082 (0.074)	0.779*** (0.247)	0.744*** (0.275)	0.121*** (0.025)	0.080* (0.044)	0.519** (0.232)	0.709*** (0.234)	0.106*** (0.041)	0.063 (0.052)	0.717*** (0.234)	0.673*** (0.253)
Participated twice or more	0.312*** (0.100)	0.261** (0.120)	0.157 (0.248)	-0.065 (0.231)	0.315*** (0.088)	0.240* (0.125)	0.227 (0.230)	0.131 (0.152)	0.316*** (0.049)	0.253*** (0.050)	0.193 (0.328)	0.349 (0.356)	0.263** (0.108)	0.191 (0.138)	0.251 (0.413)	0.186 (0.429)
<u>Age</u>		-0.010 (0.013)		-0.168* (0.092)		-0.011 (0.009)		-0.032 (0.091)		-0.011 (0.010)		-0.103 (0.097)		-0.018** (0.009)		0.037 (0.070)
<u>Gender</u> (Baseline: Male)		-0.067 (0.078)		-0.525 (0.460)		-0.051 (0.061)		-0.469 (0.320)		-0.080 (0.075)		0.119 (0.228)		-0.079 (0.064)		-0.004 (0.254)
<u>Higher education</u> (Baseline: No)		0.145 (0.112)		0.704*** (0.208)		0.162 (0.109)		0.064 (0.190)		0.062 (0.076)		0.363 (0.390)		0.151 (0.114)		0.004 (0.381)
<u>Field of Studies</u> (Baseline: Economics)																
Fundamental science		-0.012 (0.123)		-0.023 (0.587)		-0.018 (0.100)		-0.072 (0.502)		-0.162 (0.144)		0.536 (0.588)		0.000 (0.114)		0.210 (0.379)
Humanities		-0.056 (0.146)		-0.419 (0.302)		-0.074 (0.129)		-0.206 (0.295)		-0.102 (0.157)		-0.709* (0.362)		-0.110 (0.150)		-0.017 (0.471)
Other		-0.089 (0.107)		-0.168 (0.369)		-0.113 (0.085)		0.010 (0.459)		-0.120 (0.109)		-0.250 (0.419)		-0.134 (0.102)		-0.206 (0.414)
Observations	103	103	81	81	114	114	92	92	123	123	87	87	123	123	75	75
LR chi2	24.510	30.632	8.597	22.078	24.795	20.924	11.641	45.464	71.874	111.576	13.274	21.755	14.669	27.992	9.950	6.515
Prob > χ^2	0.000	0.000	0.014	0.001	0.000	0.001	0.003	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.007	0.259

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses. Marginal effects estimated at the means of covariates in brackets.

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

We observe a significant positive effect of experience in participating in economic

²⁷As subjects categorized as ”Economics” represent the greatest category among Students, we changed the baseline of the field of studies variable accordingly.

experiments on the Selfishness/Altruism parameter ($\hat{\alpha}_s$). Such an effect is even more pronounced among those who participated more than once in the past (it approximately doubles, though we do not have the specific impact of each additional year within this category). Although the impact of the first participation is reduced and often not significant when adding control variables, the impact of more than one participation is generally robust.

We also find a positive and very significant effect of having participated once in the past on the probability to be an efficiency-seeker ($\hat{\rho}_s > 0$), which resists the introduction of control variables. However, no significant effect is detected when looking at subjects who participated more than once. The overall evidence is consequently more mixed than for the Selfishness/Altruism tradeoff, although it may suggest a non-linear effect of past experience on the Equality/Efficiency tradeoff.

We note that both findings are compatible with age being confounded by experience in experimental decision-making in the literature, as in Kettner and Waichman (2016). Indeed, our results are consistent with concluding that inexperienced subjects are more eager to behave altruistically and be more sensitive to equality, which is similar to the pattern observed when comparing young and old subjects. Unfortunately, we are unable to distinguish the possible channels through which past experience in economic experiments may relate to distributional preferences. It could be that holding some experience in experimental procedures have some direct effect. It could also be that only subjects with specific preferences self-select into multiple participations. In any case, such findings would require further investigation.

B.5.5 Alternative measures of political preferences

The end-of-experiment questionnaire included three questions designed to measure subjects' political preferences: *i*) the proximity to one party among a list of best-known political parties, *ii*) a self-placement on an left-right axis consisting of 11 nodes, *iii*) a placement of the five biggest political parties on the same left-right axis.²⁸ Such questions are directly copied from the survey on political dynamics²⁹ run by the Centre de recherches politiques de Sciences Po (CEVIPOF).³⁰ The survey consists in several barometers over the 2014-2017 period which ask a representative sample of about 1500 French citizens about their political preferences, their views about public decisions and their reactions to economic and social changes as well as to the media agenda.

The regression analysis exposed in section 3 is based on question *ii*). The exact wording of such question is: "*In politics, people talk about left-wing and right-wing. On a 0 to 10 scale, where would you locate yourself?*". In this section, we describe how we use questions *i*) and *iii*) to run robustness checks on our results.

The first question asks subjects to select the political party to which they feel the closest out of a list composed of 12 significant french political parties. The exact wording of such question is: "*Without thinking only about the elections, please indicate from which political party you feel the closest, or, say, the least distant.*". The table B.11 summarizes the responses to that question, along with the responses from the national survey held by the CEVIPOF in 2017. Notice that the possible answers are classified from the more left-wing to the more right-wing (although the characterization of left and right is debatable, such a classification corresponds to the general opinion and to the way the survey is usually run). Interestingly, the mass of our Non-student control group is located in the center-left or do not have an opinion (which mirrors the proportion of "no opinion" in the national survey).

²⁸As previously mentioned, these questions were included only in the questionnaires for Politicians from the Field and for Non-Students, not in the questionnaires for Politicians from the Lab and for Students.

²⁹<https://www.sciencespo.fr/cevipof/fr/content/les-dynamiques-politiques-2014-2017.html>

³⁰The CEVIPOF is one leading institute in political researches in France: it aims at documenting and analyzing the drivers of French political institutions, attitudes and behavior.

Our Politicians also show such a center-left concentration, but also span a wider range of political parties. Besides, fewer Politicians report not having an opinion (which is expected, considering that taking part in the local political life is probably connected with a general interest for policy-making, if not politics). Interestingly, looking at the proportion of subjects who feel closer to one of the five biggest parties (LFI, PS, LREM, LR and RN), we see that it accounts for 59% of Politicians but only for 33% of Non-students, whereas most Non-students declare being close to the "green" party (in line with the "green landslide" that occurred at the 2020 french local elections).

Table B.11: Closest political party by groups (proportions)

	Pol(Field)	Non-student Control	Population (CEVIPOF, 2017)
Lutte Ouvrière ou NPA	0.00	0.00	0.01
Le Parti Communiste	0.00	0.00	0.01
La France Insoumise (LFI)	0.04	0.04	0.10
Nouvelle Donne	0.03	0.01	0.01
Le Parti Socialiste (PS)	0.26	0.16	0.06
Europe Ecologie - Les Verts	0.11	0.22	0.03
La République en Marche (LREM)	0.21	0.12	0.14
Le MoDem	0.04	0.04	0.02
L'Union des Démocrates et des Indépendants	0.06	0.00	0.01
Les Républicains (LR)	0.08	0.01	0.14
Debout la France	0.01	0.02	0.02
Le Rassemblement National (RN)	0.00	0.00	0.14
Autre	0.03	0.03	0.00
Sans opinion	0.13	0.30	0.31

Based on the answers to this question, we constituted a "party proximity" variable by grouping the answers based on a right/center/left taxonomy. Considering that some political parties close to the center are difficult to categorize (foremost the "green party"), we actually generated four different variables:

1. Left: up to PS. Center: green party to UDI. Right: down to LR.
2. Same as 1) but with UDI moved to the right.

3. Same as 2) but with green party moved to the left.
4. Same as 1) but with green party moved to the left.

The second question asks subjects to position each of the five most influential french political parties (LFI, PS, LREM, LR and RN) on the very same left/right scale that the one they used to position themselves. The exact wording of such question refers to the self-placement question and is: "*On the same 0 to 10 scale, where would you locate the following political parties*". Our variable (Big five proximity) classifies subjects into one of four groups depending on the distance between their self placement and these rankings: a subject was identified with a political party if it is the closer to her personal position on the scale.³¹ Furthermore, considering that there was few subjects that fell close to either LR (15) or RN (9), we decided to merge the two parties (hence identifying right-wing preoccupations in a broad sense).

The tables B.12 and B.13 report the results from regressing these 5 definitions of political preferences on both our estimated parameters. For each restriction strategy and each definition of political preferences (including the self-placement one), we run two models: one with the political preferences variable alongside the group variable, age and gender (first column) and one with the same variables plus the remaining "relevant" control variables (income and education).

Regarding the Selfishness/Altruism parameter, the results from the self-placement variable indicate an increasing concern for others when going from the left to the right (or to the center, considering the concentration of our data in the center-left). Yet such a negative effect is not clear-cut, as it is significantly different from 0 only for subjects-based restrictions. What we observe with alternative definitions of political preferences is similar: though we do observe some very significant effects in some cases, the pattern is overall not consistent throughout. Sometimes the "left" is indicated as significantly more selfish, sometimes the "right" is (and sometimes

³¹Notice that there could be ties: a subject could indeed position himself at equidistance from several parties. (say 2 when LFI is 1 and PS 3). This is indeed the case for 45 subjects out of 162. In such a case we classified subjects based on a random draw.

both, as opposed to the center). We therefore do not conclude in any specific impact of political preferences on the way subjects trade off selfishness and altruism.

As concerns the equality-efficiency tradeoff, results on the self-placement variable almost systematically highlight a positive association between right-wing orientation and efficiency-orientation. We recover such a result with our other characterizations of political preferences: the variable identifying the most left-wing values in each case is almost systematically and significantly most likely to reduce payoff inequalities than the baseline category. Such an effect resists the introduction of control variables and is sometimes even reinforced. Again, we note that it is more likely to reflect a left/center distinction than a left/right distinction, due to the composition of our subject pools. Furthermore, we also observe that the definition of the variable have an impact on the regression estimates: when the green party is classified on the left (party proximity 1 v. 4), the effect is stronger and more significant than when it is classified on the center. In particular, even the results from subjects-based restrictions are highly-significant when the greens are on the left. The position of UDI also plays a similar role, although it appears more moderate (party proximity 1 v. 2).

B.5.6 Heterogeneity analysis of estimated parameters within Politicians

We are particularly interested in six possible variables which could explain Politicians' heterogeneity in estimated parameters. For each variable, we look at the correlation (continuous variables) or the average per category (discrete variables) between the variable and each estimated parameter.

The most important variable is experience in policy-making, measured by the number of years spent holding local political duties. As already mentioned, the channels through which Politicians may differ from the remainder of the population involve either a selection effect or an experience effect (or both). Directly looking at the effect of experience is thus a primary way of testing the mechanism at play.

The second variable is the local status (mayor, vice-mayor or local councillor). Even though local policies are chosen based on a vote within the local councils so that every vote counts, the level of responsibilities as well as the leadership position differ among councils' members. One may argue that preferences' specificity are likely even more prominent among individuals who decided upon endorsing important responsibilities as politicians.

The last four variables are used to question whether some standard characteristics of the underlying population matter for the choice consistency of their representatives.³² In particular, we are interested in assessing whether the size of population or its wealth have any impact. Regarding the size of the population, we look into both the (log) number of inhabitants in the town of election and whether the subject holds a term at a higher administrative layer in addition to his local term (meaning he often has to decide for a larger pool of inhabitants than his sole election town). Regarding the wealth of the population, we consider the (log) income per inhabitant or the (log) tax capacity³³ per inhabitant in the town of election.

The table B.14 summarizes our results on the Selfishness/Altruism tradeoff. We do not observe much impact of both our categorical variables (either elected status or intermunicipal term). Similarly, the correlations with experience or tax capacity

³²Concerning town's characteristics, they are not self-declared but rather directly recovered from INSEE statistics based on Politicians' zip-code.

³³As a reminder, the tax capacity is a measure of the capacity in terms of tax resources that the town is able to extract from its territory.

Table B.14: Politicians specific characteristics and estimated Selfishness/Altruism tradeoff ($\hat{\alpha}_s$)

	HEI \geq 0.51	Average HEV \geq 0.927	H ² MI	HEV
<u>Political experience</u> *	.127	.131	.064	.15
<u>Population</u> (2019)*	-.152	-.142	.003	-.012
<u>Income per capita</u> (2019)*	.121	.076	.093	.133
<u>Tax capacity per capita</u> (2019)*	-.087	-.051	-.01	.001
<u>Elected status</u>				
Mayor	.61	.637	.543	.611
Vice-mayor	.676	.675	.589	.637
Local councillor	.584	.597	.586	.621
<u>Higher-layer term</u>				
No	.632	.647	.574	.628
Yes	.634	.631	.578	.624

Note: Entries report Pearson's correlation coefficient between the variable and the parameter for continuous variables (identified by *) or average parameter in the category for discrete variables.

HEI: Homothetic Efficiency Index. H²MI: Homothetic Houtman-Maks Index. HEV: Homothetic Efficiency Vector.

appear rather weak. In all cases the relationship is not statistically significant. However, we do find some statistically significant differences with respect to population and average income. An increase in population is related to an increase in concern for others, whereas an increase in income per capita is related to an increase in selfishness.

Concerning the Equality/Efficiency tradeoff, we do not observe much relationships (table B.15). Most findings show a small effect size and are furthermore not significant. We however mention two exceptions. First there is a slightly positive and significant effect of increased experience on the likelihood to have equality-oriented preferences when observations-based restrictions are implemented. Second, efficiency orientation slightly correlates significantly with average income when subjects-based restrictions are implemented.

Table B.15: Politicians specific characteristics and estimated Equality/Efficiency tradeoff ($\hat{\rho}_s$)

	HEI ≥ 0.51		Average HEV ≥ 0.927		H ² MI		HEV	
	$\hat{\rho}_s < 0$	$\hat{\rho}_s \geq 0$	$\hat{\rho}_s < 0$	$\hat{\rho}_s \geq 0$	$\hat{\rho}_s < 0$	$\hat{\rho}_s \geq 0$	$\hat{\rho}_s < 0$	$\hat{\rho}_s \geq 0$
<u>Political experience</u> *	12.404	12.16	12.343	11.406	12.823	10.435	12.905	10.25
<u>Population (2019)</u> *	8.087	8.345	8.055	8.259	8.163	8.07	8.074	8.167
<u>Income per capita (2019)</u> *	9.549	9.641	9.539	9.599	9.556	9.551	9.544	9.556
<u>Tax capacity per capita (2019)</u> *	6.478	6.532	6.487	6.493	6.509	6.454	6.513	6.463
<u>Elected status</u>								
Mayor	.214	.2	.221	.188	.167	.246	.213	.255
Vice-mayor	.554	.36	.471	.344	.483	.41	.492	.353
Local councillor	.232	.44	.309	.469	.35	.344	.295	.392
<u>Higher-layer term</u>								
No	.439	.44	.429	.438	.468	.468	.46	.481
Yes	.561	.56	.571	.563	.532	.532	.54	.519

Note: Entries report average variable in the parameter’s category for continuous variables (identified by *) or proportions of the variable’s categories within each parameter’s category for discrete variables.

HEI: Homothetic Efficiency Index. H²MI: Homothetic Houtman-Maks Index. HEV: Homothetic Efficiency Vector.

In order to check the robustness of such findings, we run regressions that include all three variables that revealed at least some degree of correlation. The results are exposed in the table B.16.³⁴ Overall, we find consistent results. Most importantly, we recover a negative impact of population and a positive impact of average income on $\hat{\alpha}_s$. Interestingly, the effects resist the addition of control variables, and even reinforce for population. However, the χ^2 test on the overall significance of the effects often fails to reject the null that all coefficients are zero in some models, which may cast some doubts as to the statistical relevance of the variables. We also recover some impact of experience and income per capita on $\hat{\rho}_s$, though the effects are not necessarily consistent throughout all the models and are not significant when observations-based restrictions are implemented. As a consequence, though we provide some pointers as to possibly interesting relationships, we cannot conclude robustly on the performance of the identified variables in explaining the heterogeneity in estimated parameters among Politicians. Additional research based on bigger

³⁴We remove the 5 subjects who chose "Do not want to answer" to the stated income question from the regressions due a the strict colinearity issue between the corresponding dummy variable and equality-orientation in some models. However, the results are virtually identical in the other models when such category is included.

APPENDIX B

Table B.16: Regressions of Politicians' specific characteristics on estimated parameters

	HEI				HEV ^a				H ² MI				HEV ^v				
	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	$\hat{\alpha}_s$	$\hat{\alpha}_s$	$\hat{\rho}_s > 0$	$\hat{\rho}_s > 0$	
<u>Length of service</u>	0.004 (0.005)	0.002 (0.006)	-0.013 (0.020)	0.011 (0.022)	0.005 (0.005)	0.005 (0.006)	-0.018 (0.017)	-0.004 (0.015)	0.002 (0.003)	0.002 (0.004)	-0.031* (0.017)	-0.034* (0.019)	0.005 (0.003)	0.004 (0.004)	-0.038** (0.018)	-0.032* (0.018)	
<u>Population 2019 (log)</u>	-0.056*** (0.021)	-0.075*** (0.021)	-0.135 (0.140)	-0.007 (0.161)	-0.053** (0.022)	-0.065*** (0.022)	-0.119 (0.129)	-0.042 (0.145)	-0.022 (0.016)	-0.029* (0.016)	-0.132 (0.121)	-0.127 (0.127)	-0.036** (0.015)	-0.044*** (0.014)	-0.052 (0.121)	-0.029 (0.145)	
<u>Income per capita 2019 (log)</u>	0.330** (0.155)	0.253* (0.146)	3.869*** (0.912)	4.920*** (1.037)	0.204 (0.157)	0.132 (0.161)	2.644*** (0.839)	3.500*** (0.914)	0.206 (0.146)	0.229* (0.136)	0.657 (1.043)	1.114 (0.950)	0.401*** (0.139)	0.344*** (0.131)	1.077 (0.920)	1.284 (1.064)	
<u>Age</u>		0.000 (0.004)		-0.014 (0.021)		0.002 (0.004)		-0.011 (0.017)		-0.002 (0.003)		0.011 (0.013)		0.002 (0.003)		-0.004 (0.013)	
<u>Gender</u> (Baseline: Male)		0.123 (0.077)		-1.230*** (0.379)		0.084 (0.074)		-1.055*** (0.343)		0.014 (0.060)		-0.796*** (0.293)		0.033 (0.068)		-0.703** (0.293)	
<u>Higher education</u> (Baseline: No)		-0.037 (0.057)		0.301 (0.293)		-0.008 (0.056)		0.440 (0.336)		-0.063 (0.046)		0.217 (0.283)		-0.033 (0.041)		0.466* (0.257)	
<u>Income</u> (Baseline: Between 2000 € and 3000 €)			-0.161** (0.069)		1.424** (0.575)		-0.021 (0.080)		1.041*** (0.402)		-0.014 (0.063)		0.767** (0.315)		-0.017 (0.063)		0.448 (0.319)
< 2000 €																0.176 (0.176)	
> 3000 €			0.052 (0.115)		0.193 (0.376)		0.068 (0.092)		0.237 (0.415)		0.073 (0.061)		0.266 (0.277)		0.059 (0.068)		0.272 (0.354)
Observations	83	83	80	80	103	103	100	100	134	134	119	119	134	134	110	110	
LR chi2	10.024	30.376	18.726	44.423	8.092	10.449	11.720	64.472	4.273	9.597	4.843	18.550	12.366	44.894	6.131	19.778	
Prob > χ^2	0.018	0.000	0.000	0.000	0.044	0.235	0.008	0.000	0.233	0.294	0.184	0.017	0.006	0.000	0.105	0.011	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the session level in parentheses. Marginal effects estimated at the means of covariates in brackets.

Note: HEI: Homothetic Efficiency Index ≥ 0.51 . HEV^a: average element in Homothetic Efficiency Vector ≥ 0.927 . H²MI: only observations not violating HARP. HEV^v: only observations with $h_s^i \geq 0.95$.

and possibly national samples would consequently be beneficial to look into such effects.

References

- Afriat, S. N. (1967). The construction of utility functions from expenditure data. International Economic Review, 8(1):67–77.
- Almås, I., Cappelen, A. W., and Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? Journal of Political Economy, 128(5):1753–1788.
- Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. Econometrica, 77(5):1607–1636.
- Andreoni, J., Gillen, B. J., and Harbaugh, W. T. (2013). The power of revealed preference tests: Ex-post evaluation of experimental design. Unpublished manuscript.
- Andreoni, J., Nikiforakis, N., and Stoop, J. (2017). Are the rich more selfish than the poor, or do they just have more money? a natural field experiment. Technical report, National Bureau of Economic Research.
- Bauman, Y. and Rose, E. (2011). Selection or indoctrination: Why do economics students donate less than the rest? Journal of Economic Behavior & Organization, 79(3):318–327.
- Beatty, T. K. and Crawford, I. A. (2011). How demanding is the revealed preference approach to demand? American Economic Review, 101(6):2782–95.
- Bellemare, C., Kröger, S., and Van Soest, A. (2008). Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. Econometrica, 76(4):815–839.
- Bellemare, C., Kröger, S., and van Soest, A. (2011). Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool. Journal of Economic Behavior & Organization, 78(3):349–365.
- Ben-Ner, A., Kong, F., and Putterman, L. (2004). Share and share alike? gender-pairing, personality, and cognitive ability as determinants of giving. Journal of Economic Psychology, 25(5):581–589.

- Bronars, S. G. (1987). The power of nonparametric tests of preference maximization. Econometrica: Journal of the Econometric Society, pages 693–698.
- Cappelen, A., List, J., Samek, A., and Tungodden, B. (2020). The effect of early-childhood education on social preferences. Journal of Political Economy, 128(7):000–000.
- Cappelen, A. W., Halvorsen, T., Sørensen, E. Ø., and Tungodden, B. (2017). Face-saving or fair-minded: What motivates moral behavior? Journal of the European Economic Association, 15(3):540–557.
- Cappelen, A. W., Moene, K. O., Sørensen, E. Ø., and Tungodden, B. (2013). Needs versus entitlements—an international fairness experiment. Journal of the European Economic Association, 11(3):574–598.
- Cappelen, A. W., Nygaard, K., Sørensen, E. Ø., and Tungodden, B. (2015). Social preferences in the lab: A comparison of students and a representative population. The Scandinavian Journal of Economics, 117(4):1306–1326.
- Croson, R. and Gneezy, U. (2009). Gender differences in preferences. Journal of Economic Literature, 47(2):448–74.
- Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Economic Theory, 33(1):67–80.
- Dean, M. and Martin, D. (2016). Measuring rationality with the minimum cost of revealed preference violations. Review of Economics and Statistics, 98(3):524–534.
- Dufwenberg, M. and Muren, A. (2006). Gender composition in teams. Journal of Economic Behavior & Organization, 61(1):50–54.
- Engel, C. (2011). Dictator games: A meta study. Experimental Economics, 14(4):583–610.
- Epps, T. and Singleton, K. J. (1986). An omnibus test for the two-sample problem using the empirical characteristic function. Journal of Statistical Computation and Simulation, 26(3-4):177–203.

- Esarey, J., Salmon, T. C., and Barrilleaux, C. (2012). What motivates political preferences? self-interest, ideology, and fairness in a laboratory democracy. Economic Inquiry, 50(3):604–624.
- Fehr, E., Naef, M., and Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. American Economic Review, 96(5):1912–1917.
- Fisman, R., Jakiela, P., and Kariv, S. (2015a). How did distributional preferences change during the great recession? Journal of Public Economics, 128:84–95.
- Fisman, R., Jakiela, P., and Kariv, S. (2017). Distributional preferences and political behavior. Journal of Public Economics, 155:1–10.
- Fisman, R., Jakiela, P., Kariv, S., and Markovits, D. (2015b). The distributional preferences of an elite. Science, 349(6254):aab0096.
- Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. American Economic Review, 97(5):1858–1876.
- Frank, R. H., Gilovich, T., and Regan, D. T. (1993). Does studying economics inhibit cooperation? Journal of Economic Perspectives, 7(2):159–171.
- Frigau, L., Medda, T., and Pelligra, V. (2019). From the field to the lab. an experiment on the representativeness of standard laboratory subjects. Journal of Behavioral and Experimental Economics, 78:160–169.
- Gross, J. and Kaiser, D. (1996). Two simple algorithms for generating a subset of data consistent with warp and other binary relations. Journal of Business & Economic Statistics, 14(2):251–255.
- Heufer, J. (2013). Testing revealed preferences for homotheticity with two-good experiments. Experimental Economics, 16(1):114–124.
- Heufer, J. and Hjertstrand, P. (2015). Consistent subsets: Computationally feasible methods to compute the houtman–maks-index. Economics Letters, 128:87–89.
- Heufer, J. and Hjertstrand, P. (2019). Homothetic efficiency: Theory and applications. Journal of Business & Economic Statistics, 37(2):235–247.

- Houtman, M. and Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. Kwantitatieve methoden, 19(1):89–104.
- Jakiela, P., Miguel, E., and Te Velde, V. L. (2015). You’ve earned it: estimating the impact of human capital on social preferences. Experimental Economics, 18(3):385–407.
- Kerschbamer, R. and Müller, D. (2020). Social preferences and political attitudes: An online experiment on a large heterogeneous sample. Journal of Public Economics, 182:104076.
- Kettner, S. E. and Waichman, I. (2016). Old age and prosocial behavior: Social preferences or experimental confounds? Journal of Economic Psychology, 53:118–130.
- Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. Journal of Econometric Methods, 1(1):23–41.
- Müller, D. (2019). The anatomy of distributional preferences with group identity. Journal of Economic Behavior & Organization, 166:785–807.
- Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., and Keltner, D. (2010). Having less, giving more: the influence of social class on prosocial behavior. Journal of Personality and Social Psychology, 99(5):771.
- Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., and Keltner, D. (2012). Higher social class predicts increased unethical behavior. Proceedings of the National Academy of Sciences, 109(11):4086–4091.
- Saad, G. and Gill, T. (2001). The effects of a recipient’s gender in a modified dictator game. Applied Economics Letters, 8(7):463–466.
- Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). Specification curve analysis. Nature Human Behaviour, 4(11):1208–1214.
- Trautmann, S. T., Van De Kuilen, G., and Zeckhauser, R. J. (2013). Social class and (un) ethical behavior: A framework, with evidence from a large population sample. Perspectives on Psychological Science, 8(5):487–497.

- Varian, H. R. (1982). The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society, pages 945–973.
- Varian, H. R. (1983). Non-parametric tests of consumer behaviour. The Review of Economic Studies, 50(1):99–110.
- Varian, H. R. (1990). Goodness-of-fit in optimizing models. Journal of Econometrics, 46(1-2):125–140.
- Varian, H. R. (1993). Goodness-of-fit for revealed preference tests. University of michigan ann arbor, Department of Economics.

Chapter 3

Preferences matter! Political Responses to the COVID-19 and Population’s Preferences¹

Contents

1	Introduction	294
2	Data and measures	296
3	Results	300
4	Concluding remarks	310
	References	312

¹This chapter is an extended version of a working paper co-authored with Etienne Dagorn & Martina Dattilo.

"In short, it appears that policy-makers, if they wish to forecast the response of citizens, must take the latter into their confidence. This conclusion, if ill-suited to current econometric practice, seems to accord well with a preference for democratic decision-making."

Robert Lucas, *Econometric Policy Evaluation: a Critique*,
Carnegie-Rochester Conference Series on Public Policy, 1976

1 Introduction

Human societies over the world and the governments that rule them are regularly confronted to social, economic and/or political crises. How they handle such crises through emergency policy-making has a substantial impact on both present and future citizens' well-being. Yet, the efficiency of policy responses crucially hinges on laypeople's behavior both during the crisis and in response to the implemented policies. Such behavior is informed by different types of preferences, which differ from one country to the other. These preferences also matter directly in the policy-making process as they influence how societies value different policy interventions, hence their political attractiveness. Assessing how preference measures within a population relate to political responses therefore is an important empirical question to improve our understanding of governments' action during a crisis. Yet, such a inquiry requires both a level of representativeness and an exogenous source of variation that are generally difficult to achieve in most cases.

The COVID-19 pandemic outbreak however offers an unique opportunity to investigate this question. Indeed, it was a highly-exogenous and mostly unexpected event that rapidly affected all parts of the world. Within a two-weeks period in early March 2020, most countries implemented stringent policies to contain the COVID-19's first wave and converged towards a similar response pattern (Hale et al., 2021). Yet, such a clustering hides substantial heterogeneity in both the timing and the intensity of the policies relative to the local dynamics of the epidemic. This chapter aims to explain the heterogeneity in political responses through the diversity of populations' preferences with respect to risk-taking, patience, altruism and trust. These types of preference have indeed been shown to be determinant for many different decisions and to differ widely across societies. Furthermore, the COVID-19 context renders such preferences particularly salient: decisions about possibly long-lasting

restrictions, whose results partly relied on citizens' pro-sociality, had to be rapidly made in a very uncertain environment. Combining open-access data on how governments reacted to the epidemic with measures of preferences obtained through large-scale representative surveys in many different countries a few years before the COVID-19 bursted thus allows for an accurate descriptive analysis.

There are two main channels through which population's preferences - in terms of risk-taking, patience, altruism and trust - and political responses may relate during a worldwide crisis.

On the one hand, population's preferences determine the need for policy intervention (Alfaro et al., 2020). Indeed, preferences influence the behavior undertaken by individuals during the course of their everyday life, which may impact the epidemic dynamics in several ways. First, the prevalence of the epidemic relies on the probability of exposure to the virus, hence on the frequency and intensity of social interactions which co-evolve with individual preferences. For instance, the number and strength of social ties are often proxied by interpersonal trust (Bowles and Polania-Reyes, 2012, Algan and Cahuc, 2013). Second, the population's spontaneous reaction to the virus's spread is shaped by preferences which determines the willingness to protect oneself and other people. Third, preferences also influence people's compliance with public policies, hence their effectiveness. In an uncertain environment where finding the optimally-balanced policy is crucial for populations' well-being, governments are indeed likely to pay close attention to the determinants of policies' effectiveness. Many empirical studies using various methodologies document such relationships. Interpersonal trust appears key to the individual adoption of spontaneous disease-mitigation behaviors - such as reducing mobility - and compliance with public health policies (Borgonovi and Andrieu, 2020, Brodeur et al., 2021a,b, Barrios et al., 2021, Durante et al., 2021, Algan et al., 2021). Similarly, trust in governments, or more broadly in healthcare systems, is critical to public interventions' effectiveness (Blair et al., 2017, Bargain and Aminjonov, 2020, Christensen et al., 2021, Han et al., 2021, Antinyan et al., 2021), with possibly important long-lasting effects (Lowe and Montero, 2021, Aassve et al., 2021). Risk-tolerant people are more likely to engage in behaviors that put their health at risk (Anderson and Mellor, 2008, Sutter et al., 2013) and less likely to take-up medicines (Simon-Tuval et al., 2018). Patient people are more likely to adhere to physical activity advice (Christensen-Szalanski and Northcraft, 1985, Chapman and Coups, 1999)

and medication (Brandt and Dickinson, 2013, Van Der Pol et al., 2017, Haushofer et al., 2019, Barron et al., 2020) or to get tested (Bradford et al., 2010). Similar patterns have also been observed during the COVID-19 crisis (Alfaro et al., 2020, Müller and Rau, 2021). Pro-social individuals are more inclined to spontaneously adopt behaviors that improve collective health, such as physical distancing, staying at home, or wearing masks (Campos-Mercade et al., 2021).

On the other hand, population's preferences interact with the demand for policy intervention, to which governments are likely sensitive. For instance, risk-averse populations would ask governments to insure them against the risk of being infected, while altruistic populations would positively value policies that protect others by decreasing contagion rates. At the same time, imposing many restrictions may also prove counter-productive if it runs against populations' demands. Enforcement indeed represents a signal of distrust on the part of the government which ultimately leads to crowding out initial health-improving motivations (Falk and Kosfeld, 2006, Sliwka, 2007, Schmelz, 2021). In other words, a high interpersonal trust may be perceived as sufficiently efficient on its own to ensure compliance with restrictions (Bowles and Polania-Reyes, 2012), and ultimately translate into a low support for stringent policies (Algan et al., 2021).

The interplay between political responses to the COVID crisis and populations' preferences therefore appears *a priori* extremely complex and thus deserves an empirical investigation. The present study offers a descriptive analysis on the relationship between populations' preferences *ex-ante* and national political responses to a worldwide crisis based on observations from a very large number of countries.²

2 Data and measures

Analyzing the behavioral determinants of political responses to an epidemic requires *i*) an international coverage of national political responses, *ii*) between and within countries variations in preferences of countries' population prior to the epidemic.

²Such an analysis does not aim at any claim for causality. In particular, it is not possible to disentangle the possible channels through which preferences may (or not) relate to political responses. Our objective is rather to benefit from a very specific situation which allows for an almost comprehensive description of governments' reactions over the world during an important crisis to bring out some key facts.

Indeed, our analysis combines daily information on COVID-related policies around the globe with pre-pandemic large-scale international surveys on preferences.

2.1 Political responses

The information related to the spread of the SARS-COV-2 and governments' political responses starting 01/01/2020 come from the Oxford COVID-19 Government Response Tracker (OxCGRT, Hale et al., 2021), partly based on COVID-19 Data Repository from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (Dong et al., 2020). The OxCGRT reports daily political actions among 19 national-level policy indicators in more than 180 countries (see C.1.1 for details). From these 19 policy indicators, we retain 9 indicators that entail legal restrictions on individual behavior.³ Among these indicators, we restrict our attention to "restrictive policies", *i.e.* mandatory policies (hence discarding recommendations). Our perspective is indeed to focus on policies that had a direct impact on population's daily life by implementing substantial restrictions on individuals. In addition, narrowing the analysis to mandatory policies also reduce measurement errors and mitigate comparability issues that arise in international comparisons.

The present study focuses on two dimensions of governments' political responses to the epidemic: the *Responsiveness* and the *Stringency* of the responses (see figures C.3, C.4 and C.5 in the section C.2 from appendix C for worldwide distributions of our main variables of interest).

Responsiveness measures how quickly a given country has responded to the spread of the virus and it is approximated by minus the logarithm of the number of cases recorded in the country on the day it implemented its first restrictive policy. The higher the number of recorded cases, the less responsive its government is.

Stringency measures how strong a given country has responded to the spread of the virus and is measured by a Stringency Index, whose computation is based on Hale et al. (2021)'s methodology. Specifically, at a given date t , each of the nine restrictive

³We discarded economic policies or non-binding health policies. The nine indicators are school closing, workplace closing, canceling of public events, restrictions on gathering, closed public transportation, stay-at-home requirements, travel bans (both internally and internationally) and facial coverings. The figure C.2 in the section C.2 from appendix C depicts the distribution of the first policy being implemented.

policies is attributed a normalized score from 0 to 100 depending on its intensity and the index is calculated as the average of these scores. Noticeably, the index equally weighs the nine restrictive policies but distinguishes between national and targeted policies by halving the scores on the latter. We compute the Stringency Index for every day over the year following the day D_0 on which the country implemented its first restrictive policy.⁴

2.2 Preferences

We consider four types of preferences: willingness to take risks, patience (preferences over the inter-temporal timing of the rewards), altruism (a concern for others), and trust⁵ (the belief that others are trustworthy). Our empirical strategy builds on preference measures from different sources in order to obtain a dataset with extensive external validity due to *i*) the inclusion of a large number of countries with varying development levels from all continents and *ii*) representative samples from the national populations aged 15+ obtained through probability-based sampling. In addition, our measures of preferences were collected before the COVID-19 outbreak, implying that they are not biased by the sanitary crisis nor by the way governments handled it.

Our primary measure of preferences relies on the Global Preferences Survey (GPS, Falk et al., 2018). Embedded in the Gallup World Poll 2012 - a yearly international survey covering 76 countries for this year - the GPS provides several preferences measures, including measures of risk and time preferences, trust and altruism. As a complementary measures of preferences, we also extract information from the 6th wave of the World Values Survey (WVS, Inglehart et al., 2014), which was collected in 60 countries from 2010 to 2014.⁶ The appendix C.1.5 compares the GPS and the WVS in terms of surveyed countries. In total, our analysis comprises

⁴We consider only the first year in order to focus on pre-vaccine policies, considering that the vaccination policy largely turned the tide. Furthermore, the end of 2020 is also marked by the emergence of new COVID variants (alpha, bêta, gamma) which modified the worldwide dynamics of the virus's spread.

⁵Unless specified otherwise, we will henceforth take "trust" as meaning "interpersonal trust".

⁶In both the GPS and the WVS 6th wave, we removed China from the analysis due to missing data regarding the first policy implemented in this country (our dataset starts in January 2020, whereas China was hit and reacted in 2019). Ultimately, the number of countries included in our analysis is respectively 75 and 59.

109 different countries. For each measure of preferences, we aggregate observations at the national level (average) and standardize the result at the international level (z-score).⁷

The Global Preferences Survey is a cross-validated high-quality survey which is based on a innovative method to extract preferences from declarative questions. Specifically, it relies on an *ex-ante* validation procedure which consists in a laboratory experiment where participants answered declarative items and played standard incentivized experimental games (Falk et al., 2016). The combination of declarative items that best predicted observed behavior in the games was then included in the international survey. This approach guarantees a large internal validity, through its link with lab-environment behaviors, and substantial external validity, thanks to its representative-survey nature.

The World Values Survey (WVS) is a cross-national comparative survey providing time-series indicators on human beliefs, values and motivations. Using the WVS allows to extend our analysis along three dimensions. First, it provides alternative measures of preferences. Second, it covers a different set of countries as compared to the GPS (see appendix C.1.5), thus widening the overall geographical area under scrutiny. Third, the WVS incorporates several components of trust regarding both interpersonal and institutional trust (*e.g.* towards relatives, strangers, governments, World Health Organization, science *etc.*). We are thus able to refine our results on interpersonal trust⁸ and to consider trust in governments as a supplementary variable of interest.

⁷The definition of preferences for each data source is summarized in appendices C.1.2 and C.1.3. As an additional robustness check, we also extract information from the joint WVS ((7th wave) Haerpfer et al., 2020)/EVS ((European Values Survey) EVS, 2020, 2021) dataset (EVS/WVS, 2021), which covers 79 countries. The main advantage of the joint dataset as compared to the WVS 6th wave is the bigger number of surveyed countries, hence stronger external validity. However, the WVS/EVS dataset has a couple of drawbacks: i) risk preferences are missing, and ii) some countries are surveyed only after the beginning of the epidemic.

⁸Specifically, we construct two variables from the survey questions on trust in various groups of people (from family to foreigners). "Narrow" trust is the score related to "strangers" and is likely closest to the GPS measure. We primarily rely on such measure in our main analyses. "Global" trust adds up the scores from all the survey questions on trust in others and constitutes a robustness check on our results.

3 Results

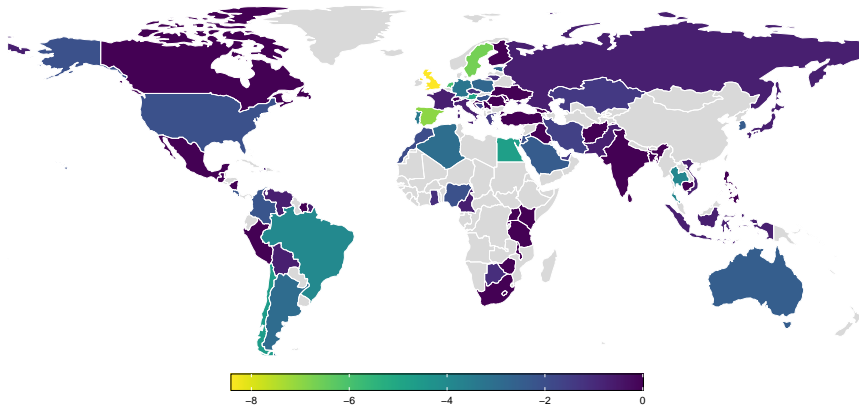
We compute country-level political responses in terms of both *Responsiveness* and *Stringency*. We approximate *Responsiveness* by minus the log number of recorded cases by the time of the first mandatory policy being implemented. We compute *Stringency* based on the methodology proposed by Hale et al. (2021) which provides a granular measure of the intensity of the overall political response for each country. The index ranges from 0 (slightly-coercive response) to 100 (highly-coercive response).

The figure 3.1 depicts substantial heterogeneity over the world in terms of both *Responsiveness* (figure 3.1a) and *Stringency* (figure 3.1b). Such heterogeneity does not appear geographically concentrated. The United Kingdom, Spain, Sweden, the Netherlands and Austria waited the most before implementing their first restrictive policy. On the contrary, countries from all over the globe reacted as soon as they recorded their first case (and even before).⁹ With respect to *Stringency*, we again observe countries from very diverse locations that adopted policies of similar intensity. Kenya, Mexico, Peru, Serbia and South Africa implemented the most stringent policies on the third day following their first restrictive policy, whereas Bangladesh, Cameroon, Canada, Croatia and Japan implemented the least stringent policy over the same time window. In other words, the international spread of the virus does not strike as the main driver of cross-countries differences in COVID-related restrictions. Scrutinizing country-specific characteristics is thus likely to help in explaining political responses to the epidemic.¹⁰

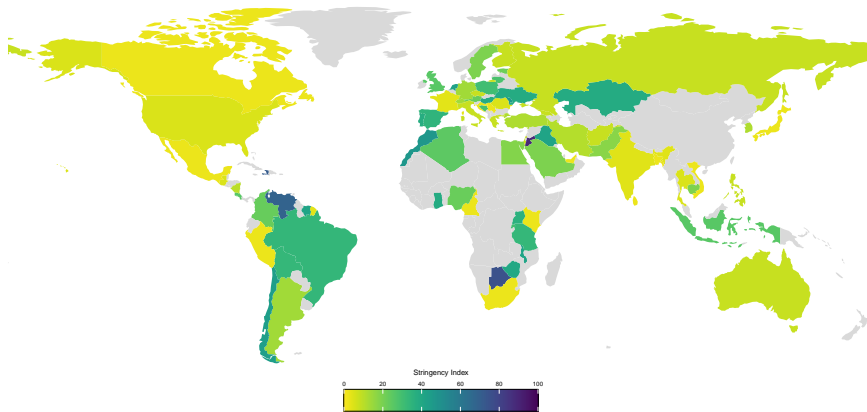
⁹E.g. Egypt, Chile, Brazil, Thailand or Portugal.

¹⁰The appendix C.2 displays similar maps with respect to alternative definitions of *Responsiveness* and various time spans concerning *Stringency* (see figures C.3, C.4 and C.5).

Figure 3.1: Worldwide political responses



(a) *Responsiveness*



(b) *Stringency at D_3*

Note: *Responsiveness* is computed as minus the log number of recorded cases by the time of the first mandatory policy being implemented. For instance, a value of -4 corresponds to approximately 55 recorded cases. *Stringency* is an index ranging from 0 (slightly-coercive response) to 100 (highly-coercive response). It is measured every day following the day where the first restrictive policy is implemented (e.g., D_3 is D_0 plus 3, hence the third day).

3.1 Governments' *Responsiveness*

The figure 3.2 reports the estimated coefficients and corresponding 95% confidence intervals in OLS regressions of governments' *Responsiveness* on nationally-aggregated preferences using either data from the GPS data (left panel) or from the WVS 6th wave (right panel).¹¹ For each preference, we plot the results from a model with all the preference under scrutiny only (purple triangle markers) and from a model which adds country-level control variables (blue circle markers).¹²

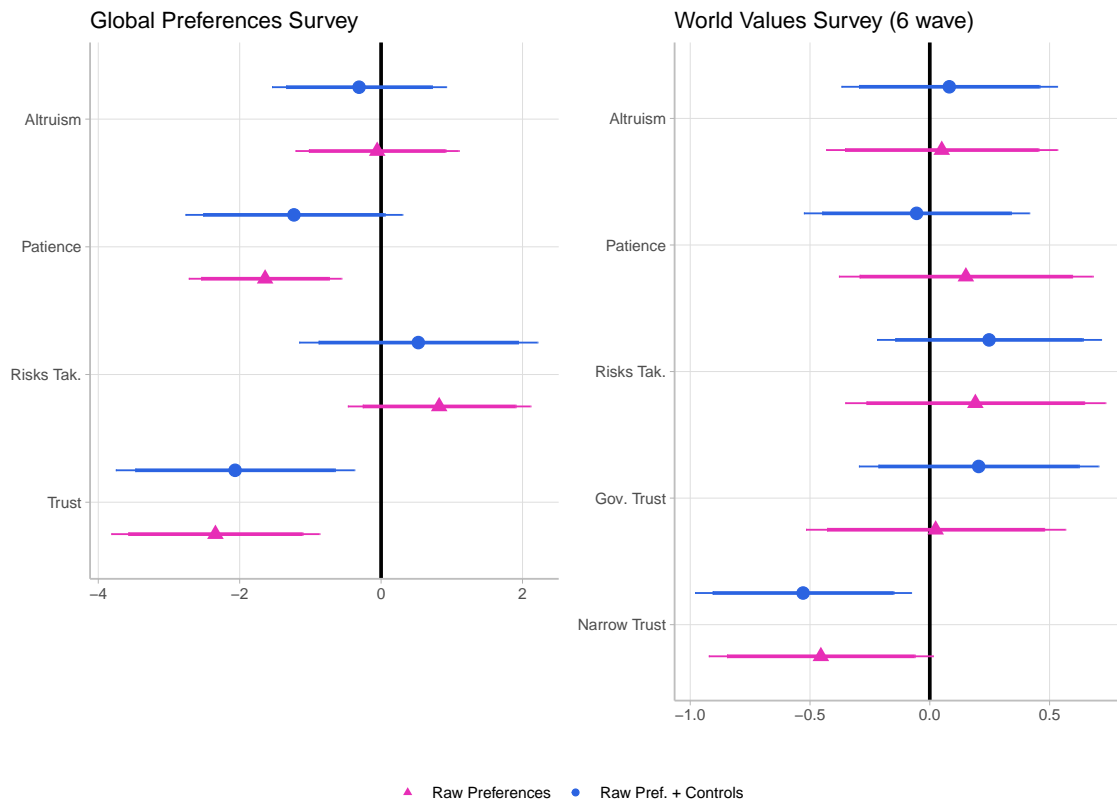
Most noticeably, we observe a negative and highly-significant correlation between *Responsiveness* and trust in the GPS. A one standard deviation (0.271) in the level of trust within the population corresponds to a 234.5% increase in the number of cases when the first restrictive policy is implemented ($p < 0.01$). Such a relationship remains remarkably stable in both its magnitude and statistical significance upon introducing control variables. In other words, governments of populations with higher levels of trust appear more willing to wait before implementing COVID-related policies. We also observe a similar though slightly weaker relationship with patience ($\beta = -1.64$, $p < 0.05$). Yet, the coefficient on patience is reduced by approximately one third and is no longer significant once country-level characteristics are accounted for.¹³ At the same time, the willingness to take risk and altruism do not strongly relate to *Responsiveness*: both coefficients are small and we fail to reject the null hypothesis that they are not different from 0. Finally, we observe that the model with all preferences account for approximately 25% of the variance in

¹¹The precise OLS specification is exposed in appendix C.3.

¹²The control variables are population's *density*, *median age*, $\log(\text{GDP per capita})$ and *worldwide COVID spread*. The first three variables aim at controlling for cross-countries differences in the epidemic's strength due to populations' demographics (see e.g. Jinjarak et al., 2020). The worldwide COVID spread is the difference in days between the first worldwide case in our dataset and the first case within the country. It aims at accounting for the non-random spatial spread of the virus, which may have caused lately-impacted countries to benefit from first-impacted countries' experience in dealing with the epidemic. See C.1.4 for more details on control variables.

¹³Patience is indeed positively and highly-significantly correlated with GDP per capita (see the correlation table in appendix C.2.2). In addition, we acknowledge that the negative relationship between patience and *Responsiveness* may seem at odds with a simple microeconomic reasoning. Indeed, a patient population is expected to care about future costs in terms of health, which would translate quite logically into a higher demand for policy intervention. Yet, such a reasoning neglects that patient populations may spontaneously adopt a disease-mitigation reaction for the exact same reason, which would reduce the need for policy intervention. This could have been anticipated by governments, so that the sign of the relationship between patience and *Responsiveness* is ultimately not straightforward. Again, our analysis is descriptive and does not intend to provide a causality stance.

Figure 3.2: *Responsiveness* and preferences



Note: Each dot corresponds to the estimated coefficient associated with a given preference measure in OLS regressions of governments' *Responsiveness* on nationally-aggregated preferences. The bars correspond to 90% (shorter) and 95% (larger) confidence intervals, computed with robust standard errors. If one bar overlaps 0, the corresponding coefficient is not statistically significant at the given confidence level. *Responsiveness* is minus the log number of cases on the date when the first restricted policy is implemented. The regressions include either only the preference measures ("raw preferences") or control for *population density*, *median age*, $\log(GDP)_{pc}$ and *worldwide COVID spread*. The left panel displays countries surveyed in the GPS (75 countries), the right panel those taking part of the 6th wave of the WVS (59 countries).

government's *Responsiveness* (and up to 30% when controls are added).¹⁴ As a consequence, populations' economic preferences (and foremost trust) appear strongly related to governments' eagerness to address the epidemic.

We recover the negative relationship between trust and *Responsiveness* in the WVS 6th wave, although it is slightly reduced in terms of both magnitude and significance, ($\beta = -0.53$, $p < 0.1$). Again, such correlation resists the introduction of

¹⁴See C.4.2 for full regressions tables.

country-level characteristics.¹⁵ Similarly, all other correlations between *Responsiveness* and preferences based on the WVS data are small and non-significant, including trust in the government.¹⁶ As a complementary robustness check, we run the same analysis on the WVS/EVS dataset (78 countries) in appendix C.4.1. The results on trust are closer to what we observed with the GPS data: a highly-significant negative relationship which resists the introduction of control variables.¹⁷ Similarly, the relationships with altruism and trust in the government are virtually unchanged (small and non-significant). However, we also uncover a positive and significant correlation between patience and *Responsiveness*, contrary to what we find in the other two datasets.

In appendix C.4.4, we run similar analyses by considering continents separately¹⁸ and find that coefficients' signs are relatively consistent throughout.¹⁹ Such a finding suggest that our relationships are not driven by some geographical imbalances, such as the unequal distribution of wealth at the international level or specific cultural traits.

Finally, we check the robustness of our results to a change in the definition of governments' *Responsiveness* (see appendix C.4.1). First, we consider taking minus the number of days between the first recorded case and the first restrictive policy in a country (definition R_2). Second, we consider substituting the number of cases per million inhabitants for the absolute number of cases in our main definition (definition R_3). In both cases, all our results stand in our primary database of interest (GPS).²⁰ We also observe consistent results in our other datasets based on definition R_2 . The results from the WVS 6th wave and the WVS/EVS based on definition R_3 are

¹⁵We exploit the richness of the WVS with respect to trust measures to construct a "global trust" variable by adding up the scores from the survey questions on trust in various groups of people (from family to foreigners). The relationship between *Responsiveness* and trust based on such variable is exactly the same as with the narrow definition of trust (see appendix C.4).

¹⁶Our results remain qualitatively similar using either the GPS or the WVS 6th wave when omitting possible outliers (*i.e.* responsiveness above 6) or when implementing a Tobit specification to account for data concentration at 0.

¹⁷We obtain even stronger results when restricting the analysis to the 64 countries surveyed before 2020.

¹⁸Due to small sample size issues, we consider only Europe, Asia and Africa as continents.

¹⁹Though our results do not systematically stand for all continents taken separately, we do not observe significant contradictory relationships across continents.

²⁰A one standard deviation in the level of trust is associated with a two-weeks longer wait or with a 340% increase in the number of cases per million inhabitants before implementing the first restrictive policies.

however more mixed. In any case, the sign of the coefficient associated with either versions of trust is systematically negative. However, none of the relationships are significant in the WVS 6th wave dataset. We observe some strong correlations in the WVS/EVS dataset, which vanish when country-level controls are added.²¹ Overall, our main results thus appear robust to changing the definition of *Responsiveness*.

3.2 Stringency of the political responses

Our analysis about the *Stringency* of the political responses builds upon a fixed effect panel regression framework with the Stringency Index computed at each dates over a year as dependent variable.²² The independent variables of interest are interaction terms between each of the four preference measures and the day dummies D_t : $\text{Pref} \times D_t$, with $t \in [1, 365]$. Such a process enables to analyse the evolution of the relationship between economic preferences and policies' *Stringency* in a year since the day when the country implemented its first restrictive policy in reaction to the epidemic.

The results are depicted in the figures 3.3 and 3.4, each of which consists of several sub-figures. Each sub-figure provides information about the strength and statistical significance of the relationship between the Stringency Index and one preference measure over time. The position of each dot with respect to the y-axis gives the sign and magnitude of the coefficient associated with the interacted variable for the corresponding preference measures at the given date t . The color of each dot informs about the p-value calculated in the estimation.²³ The lighter the color, the closest the p-value is from zero, hence the more statistically significant the corresponding relationship is.²⁴ Our goal is not to obtain precise estimates on the impact of preferences on policies' intensity, but rather to assess whether we detect some clear correlation patterns over time. In addition, such a graphical approach bypasses issues with how regression results are usually exposed (model selection,

²¹We also recover some positive and robust relationship between patience and *Responsiveness* with the WVS/EVS data.

²²The precise specification is exposed in appendix C.3.

²²We also systematically control for the within-country epidemic dynamics by including the (log) number of recorded cases at date t in the regression.

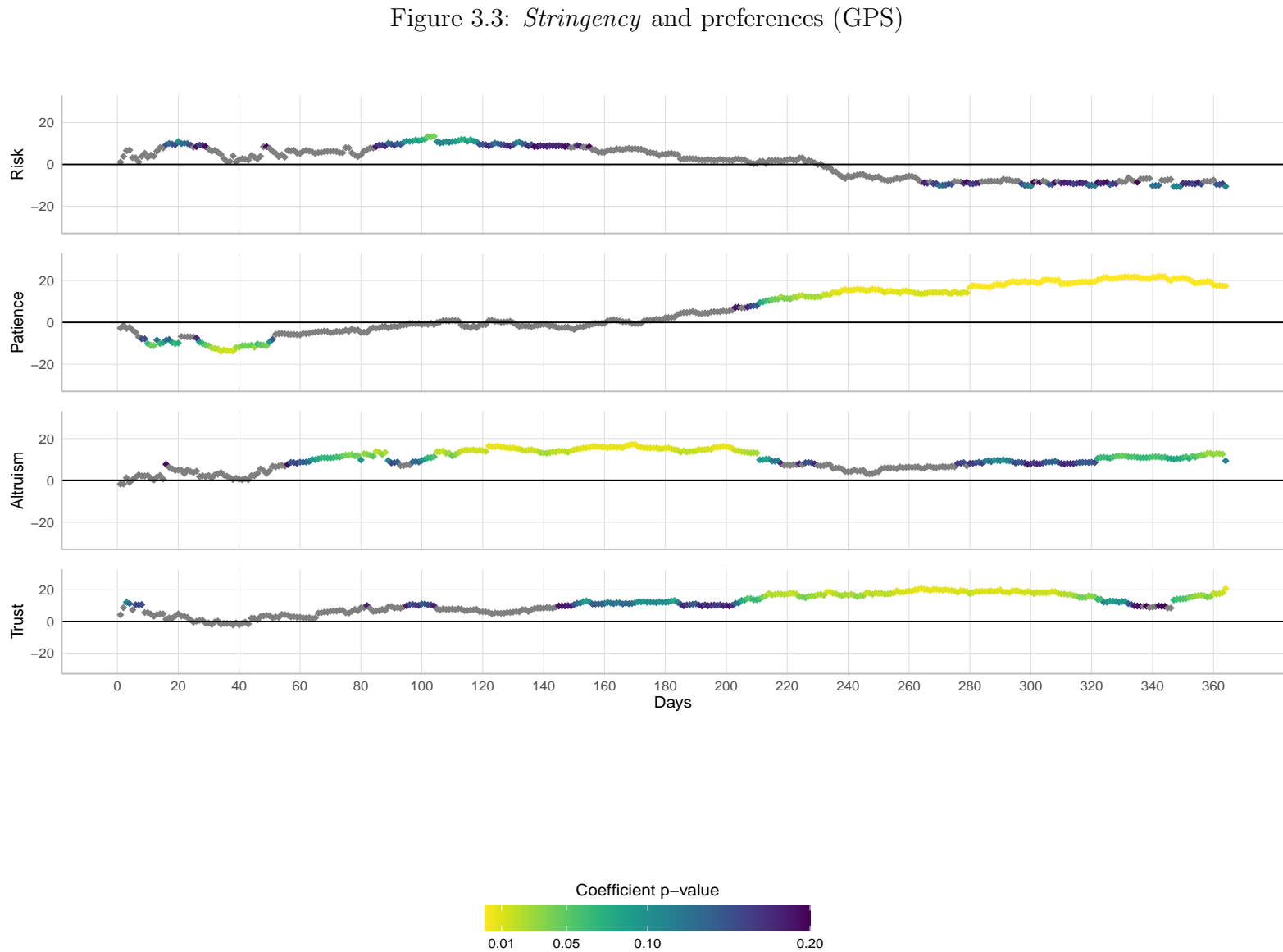
²³The computation of p-values relies on heteroskedasticity-robust standard errors.

²⁴Globally, the yellow/green/blue colors identify the 1/5/10% usual thresholds on statistical significance.

p-hacking, see e.g. Baker, 2016): we neither select a specific (arbitrary) date nor report only the standard thresholds on statistical significance.

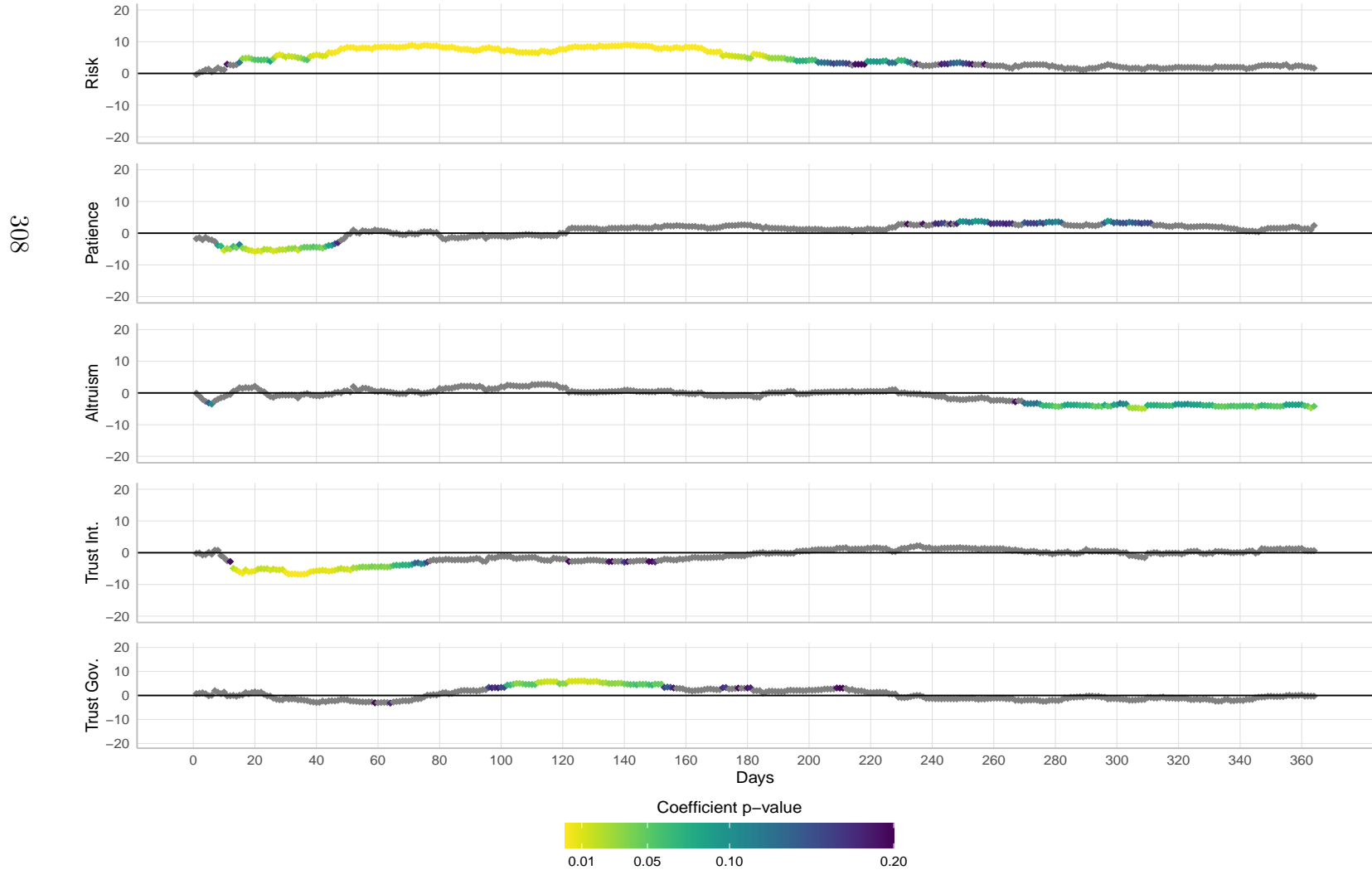
The figure 3.3 plots the regressions for countries surveyed in the GPS. Most noticeably, we find differentiated patterns over time for all preferences. Overall, altruism and trust are similarly related to *Stringency*: governments of altruistic (resp. trusting) populations implement stricter restrictions than government of self-concerned (resp. defiant) populations. In both cases, a one standard deviation in the level of the preference measure increases the Stringency Index by up to 20 points. However, the statistical significance of such patterns varies importantly with time. In particular, the relationships are not significant in the early weeks following the day when the first restrictive policy is implemented (D_0). Interestingly, the evolution over time of the relationships' statistical significance is inversely-related between altruism and trust. Altruism is highly-significantly related to *Stringency* from three to seven months after D_0 , whereas the relationship is much less significant afterwards. The converse is observed with trust.²⁵ Concerning patience, the relationship appears dual as the coefficient increases over time. In the short-run after D_0 , governments of patient populations implement looser restrictions than governments of impatient populations. However, the relationship reverses in the long-run, and especially seven months after governments' started to address the epidemic. Finally, the relationship between populations' willingness to take risk and the Stringency Index is weaker. We find a weakly-significant positive relationship over two months starting three months after D_0 , meaning that risk-loving populations implement stricter restrictions than risk-averse population during this period.

²⁵Since we measure *Stringency* starting on the day of the first restrictive policy being implemented in the country, the calendar dates for each t are different across countries. Overall, D_0 ranges from mid-January to early-March. D_{90} thus falls between mid-April and early-June, hence at a time when the pressure due to the pandemic started to decline overall. Similarly, D_{210} falls between mid-August and early-October, hence at a time when the worldwide number of cases started to soar rapidly again.



Note: The figure illustrates 364 coefficients from a panel regression of economic preferences on the Stringency Index at day D_t ($t \in [1, 365]$) along with several control variables. Each sub-graph displays estimates for each of the four preference measures obtained from the 75 countries included in the GPS. Each regression includes all preference variables (*e.g.* risk-taking, patience, altruism or trust) interacted with the day dummies D_t as well as the (log) number of cases at date t , the day dummies and country-level fixed effects. The graph for each preference summarizes the information obtained from the regressions about the corresponding preference. The dot for date t conveys two pieces of information: the position indicates both the sign and magnitude of the coefficient associated with $\text{Pref} \times D_t$ while the color indicates the statistical significance of the coefficient (p-value from two-sided t-test with heteroskedasticity-robust standard errors). The lighter the color, the closer the p-value is from zero. A grey color indicates a p-value greater than 0.2.

Figure 3.4: *Stringency* and preferences (WVS)



808

CHAPTER 3

Note: The figure illustrates 364 coefficients from a panel regression of economic preferences on the Stringency Index at day D_t ($t \in [1, 365]$) along with several control variables. Each sub-graph displays estimates for five preference measures obtained from the 59 countries surveyed in the 6th wave of the WVS. Each regression includes all preference variables (*e.g.* risk-taking, patience, altruism or trust) interacted with the day dummies D_t as well as the (log) number of cases at date t , the day dummies and country-level fixed effects. The graph for each preference summarizes the information obtained from the regressions about the corresponding preference. The dot for date t conveys two pieces of information: the position indicates both the sign and magnitude of the coefficient associated with $\text{Pref} \times D_t$ while the color indicates the statistical significance of the coefficient (p-value from two-sided t-test with heteroskedasticity-robust standard errors). The lighter the color, the closer the p-value is from zero. A grey color indicates a p-value greater than 0.2.

The figure 3.4 plots the regressions for countries surveyed in the WVS 6th wave. Globally, we observe that both the magnitude and statistical significance of the coefficients are lower than with the GPS. The overall patterns for risk and time preferences are similar with previous observations, with some differences. In particular, the positive relationship with the willingness to take risks is highly significant, whereas the positive relationship with patience in the long-run is much less significant. At the same time, the relationships between *Stringency* and either altruism or trust appear at first sight at odd with our findings on the GPS. Both preference measures relate to the Stringency Index in opposite ways with respect to previous observations. The correlation with altruism is not significant until nine months after D_0 and is significantly negative afterwards. The correlation with trust is significantly negative in the short-run and non-significant otherwise. However, comparing the whole patterns over time we observe that translating the GPS patterns downwards approximately give the WVS patterns. Consequently, though the interpretations slightly differ, the results over time are consistent across the two datasets. The slight discrepancies may thus originate from either the differing size or composition of the countries' pool across datasets.²⁶ Running the analysis on the EVS/WVS consolidates this view of consistent results overall: we recover both positive correlations with altruism and trust for the very same periods as in the GPS, while the findings on patience are also globally coherent (see figure C.11 in appendix C.4.3). Additionally, we correlate the Stringency Index with trust in the government in the WVS 6th wave and the EVS/WVS. Again, we observe a consistent pattern over time across the two datasets, although the period where the coefficients are significantly different from zero are different and yield rather opposite interpretations. Specifically, the correlation is negative in the early weeks (EVS/WVS) and positive during sixty days starting about three months after D_0 .

Finally, we replicate our analysis on each continent separately to assess whether our findings crucially rely on some geographical clusters (see appendix C.4.4). We observe an important variability both within and between continents. Specifically, the relationships for Europe are the closest to the aggregate relationships in the GPS

²⁶Interestingly, substituting the narrow with the global version of trust yields a pattern closer to our findings on the GPS, while the patterns on the other preference measures are virtually unchanged (see figure C.10 in appendix C.4.3). We take that as additional evidence that the results are consistent across the two datasets.

whereas the relationships for Asia are the closest to the aggregate relationships in the WVS 6th wave.²⁷ In both cases, the findings on the other continents are less clear-cut, but we still observe consistent patterns.²⁸ That despite some heterogeneity across continents and variations in countries across datasets we still observe coherent results overall therefore supports our findings.

4 Concluding remarks

The current study documents the relationship between several key economic preferences at the national level and countries' political responses during the SARS-CoV-2 epidemic. Combining national-level data on worldwide political responses with several large-scale representative surveys on individual preferences in a total of 109 countries, we relate both governments' responsiveness and the stringency of implemented policies at the very beginning of the sanitary crisis with pre-crisis measures of populations' willingness to take risk, patience, trust and altruism. Our results indicate that nationally-aggregated preferences are associated with both the timing and the intensity of policies designed to tackle the epidemic.

Our study provides a descriptive empirical analysis on the link between populations' preferences and how governments reacted under a worldwide crisis. Obviously, such an analysis cannot make a stance for causality nor precisely pinpoint any specific mechanism possibly at play. Such an endeavour would likely have required the building and estimation of a dynamic model relating population's preferences and political responses. However, we are capable to identify clear, highly-significant and robust correlation patterns using data from most countries over the world, and despite information loss caused by aggregating information at the country level.

Our analysis points towards one important stylized fact: populations' preferences matter for policy-making under a worldwide sanitary crisis. First, countries with high levels of trust address the crisis later than countries with low levels of trust. Second, trust correlate with the stringency of implemented policies: countries with high levels of trust either implement looser restrictions in the early weeks following the first intervention, or stricter restrictions when the epidemic strikes back later.

²⁷Europe (resp. Asia) represents a third of the GPS (resp. WVS 6th wave).

²⁸We notice only one important contradictory pattern: the european countries surveyed in the WVS 6th wave show a negative correlation between *Stringency* and patience in the long-run. Otherwise, all correlations are either consistent or at worst non-significant.

Together, such findings would be consistent with the hypothesis that governments of high trust populations prioritize citizens' demands for individual responsibility at the beginning of the sanitary crisis, and then revise their policies to tackle the virus more efficiently after a learning period. The overall pattern for patience is close to the pattern for trust: responses from governments of patient populations are looser in the short-run and stricter in the long-run than the responses from governments of impatient populations. Populations willing to take risks or concerned about others are also confronted to more stringent restrictions, especially at the time countries start to withdraw part of their containment plans following the roll-back of the epidemic.

Ultimately, this study emphasizes how different societies reacted differently to a strikingly-unexpected event, which provides valuable insights for future worldwide crises. Indeed, as global changes are likely to increase the prevalence of such events, the present analysis suggests the need to build up models integrating populations' preferences as parameters that influence both individual behavior and national political responses.

References

- Aassve, A., Alfani, G., Gandolfi, F., and Le Moglie, M. (2021). Epidemics and trust: the case of the spanish flu. Health Economics, 30(4):840–857.
- Alfaro, L., Faia, E., Lamersdorf, N., and Saidi, F. (2020). Social interactions in pandemics: fear, altruism, and reciprocity. Technical report, National Bureau of Economic Research.
- Algan, Y. and Cahuc, P. (2013). Trust and growth. Annual Review of Economics, 5(1):521–549.
- Algan, Y., Cohen, D., Davoine, E., Foucault, M., and Stantcheva, S. (2021). Trust in scientists in times of pandemic: Panel evidence from 12 countries. Proceedings of the National Academy of Sciences, 118(40).
- Anderson, L. R. and Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. Journal of Health Economics, 27(5):1260–1274.
- Antinyan, A., Bassetti, T., Corazzini, L., and Pavesi, F. (2021). Trust in the health system and covid-19 treatment. Frontiers in Psychology, 12:2733.
- Baker, M. (2016). Statisticians issue warning over misuse of p values. Nature News, 531(7593):151.
- Bargain, O. and Aminjonov, U. (2020). Trust and compliance to public health policies in times of covid-19. Journal of Public Economics, 192:104316.
- Barrios, J. M., Benmelech, E., Hochberg, Y. V., Sapienza, P., and Zingales, L. (2021). Civic capital and social distancing during the covid-19 pandemic. Journal of Public Economics, 193:104310.
- Barron, K., Damgaard, M., Gravert, C. A., and Norrgren, L. (2020). Time preferences and medication adherence: A field experiment with pregnant women in south africa. Working Paper 20-29, The Center for Economic Behavior and Inequality.

- Blair, R. A., Morse, B. S., and Tsai, L. L. (2017). Public health and public trust: Survey evidence from the ebola virus disease epidemic in liberia. Social Science & Medicine, 172:89–97.
- Borgonovi, F. and Andrieu, E. (2020). Bowling together by bowling alone: Social capital and covid-19. Social Science & Medicine, 265:113501.
- Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? Journal of Economic Literature, 50(2):368–425.
- Bradford, W. D., Zoller, J., and Silvestri, G. A. (2010). Estimating the effect of individual time preferences on the use of disease screening. Southern Economic Journal, 76(4):1005–1031.
- Brandt, S. and Dickinson, B. (2013). Time and risk preferences and the use of asthma controller medication. Pediatrics, 131(4):e1204–e1210.
- Brodeur, A., Gray, D., Islam, A., and Bhuiyan, S. (2021a). A literature review of the economics of covid-19. Journal of Economic Surveys, 35(4):1007–1044.
- Brodeur, A., Grigoryeva, I., and Kattan, L. (2021b). Stay-at-home orders, social distancing, and trust. Journal of Population Economics, pages 1–34.
- Campos-Mercade, P., Meier, A. N., Schneider, F. H., and Wengström, E. (2021). Prosociality predicts health behaviors during the covid-19 pandemic. Journal of Public Economics, 195:104367.
- Chapman, G. B. and Coups, E. J. (1999). Time preferences and preventive health behavior: Acceptance of the influenza vaccine. Medical Decision Making, 19(3):307–314.
- Christensen, D., Dube, O., Haushofer, J., Siddiqi, B., and Voors, M. (2021). Building resilient health systems: Experimental evidence from sierra leone and the 2014 ebola outbreak. The Quarterly Journal of Economics, 136(2):1145–1198.
- Christensen-Szalanski, J. J. and Northcraft, G. B. (1985). Patient compliance behavior: the effects of time on patients’ values of treatment regimens. Social Science & Medicine, 21(3):263–273.

- Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. The Lancet Infectious Diseases, 20(5):533–534.
- Durante, R., Guiso, L., and Gulino, G. (2021). Asocial capital: Civic culture and social distancing during covid-19. Journal of Public Economics, 194:104342.
- EVS (2020). European values study 2017: Integrated dataset (evs 2017). GESIS Data Archive, Cologne. ZA7500 Data file Version 4.0.0, doi:10.4232/1.13560.
- EVS (2021). European values study 2017: Ukraine (evs 2017). GESIS Data Archive, Cologne. ZA7539 Data file Version 1.0.0, doi:10.4232/1.13714.
- EVS/WVS (2021). European values study and world values survey: Joint evs/wvs 2017-2021 dataset (joint evs/wvs). GESIS Data Archive, Cologne. ZA7505. Dataset Version 2.0.0, doi:10.4232/1.13737.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. The Quarterly Journal of Economics, 133(4):1645–1692.
- Falk, A., Becker, A., Dohmen, T. J., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Working paper, Netspar discussion paper.
- Falk, A. and Kosfeld, M. (2006). The hidden costs of control. American Economic Review, 96(5):1611–1630.
- Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., and others (eds.) (2020). World values survey: Round seven—country-pooled datafile version. Madrid, Spain & Vienna, Austria: JD Systems Institute & WWSA Secretariat. doi.org/10.14281/18241.13.
- Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., et al. (2021). A global panel database of pandemic policies (oxford covid-19 government response tracker). Nature Human Behaviour, 5(4):529–538.

- Han, Q., Zheng, B., Cristea, M., Agostini, M., Bélanger, J. J., Gützkow, B., Kreienkamp, J., Leander, N. P., Collaboration, P., et al. (2021). Trust in government regarding covid-19 and its associations with preventive health behaviour and prosocial behaviour during the pandemic: a cross-sectional and longitudinal study. *Psychological Medicine*, pages 1–11.
- Haushofer, J., John, A., and Orkin, K. (2019). Can simple psychological interventions increase preventive health investment? Technical report, National Bureau of Economic Research.
- Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., and others (eds.) (2014). World values survey: Round six–country-pooled datafile version. <https://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>. Madrid: JD Systems Institute.
- Jinjarak, Y., Ahmed, R., Nair-Desai, S., Xin, W., and Aizenman, J. (2020). Accounting for global covid-19 diffusion patterns, january–april 2020. *Economics of Disasters and Climate Change*, 4(3):515–559.
- Lowes, S. and Montero, E. (2021). The legacy of colonial medicine in central africa. *American Economic Review*, 111(4):1284–1314.
- Müller, S. and Rau, H. A. (2021). Economic preferences and compliance in the social stress test of the covid-19 crisis. *Journal of Public Economics*, 194:104322.
- Schmelz, K. (2021). Enforcement may crowd out voluntary support for covid-19 policies, especially where trust in government is weak and in a liberal society. *Proceedings of the National Academy of Sciences*, 118(1).
- Simon-Tuval, T., Shmueli, A., and Harman-Boehm, I. (2018). Adherence of patients with type 2 diabetes mellitus to medications: the role of risk preferences. *Current Medical Research and Opinion*, 34(2):345–351.
- Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, 97(3):999–1012.
- Sutter, M., Kocher, M. G., Glätzle-Rützler, D., and Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents’ field behavior. *American Economic Review*, 103(1):510–31.

CHAPTER 3

Van Der Pol, M., Hennessy, D., and Manns, B. (2017). The role of time and risk preferences in adherence to physician advice on health behavior change. The European Journal of Health Economics, 18(3):373–386.

Appendix C

Preferences matter! Political Responses to the COVID-19 and Population's Preferences

Contents

C.1	Data Sources	318
C.2	Descriptive statistics	334
C.3	Model specification	342
C.4	Robustness checks	344
	References	359

C.1 Data Sources

C.1.1 The Oxford COVID-19 Government Response Tracker (OxCGRT)

The OxCGRT project is one leading source of information regarding the political response to COVID-19. Relying on a meticulous work of more than 400 volunteers, it reports publicly-available information regarding 19 indicators of government responses on a daily basis in more than 180 countries. Hale et al. (2021) exposes the data collection process, details the pros and cons of using such dataset and compares it to alternative data sources of political responses to the COVID-19 crisis.

The indicators are classified into three categories: *i*) containment and closure, *ii*) economic response and *iii*) health systems. The "containment and closure" category includes school closing, workplace closing, canceling of public events, restrictions on gathering, closed public transportation, stay-at-home requirements, and travel bans (both internally and internationally). The "economic response" category includes income support, debt/contract relief for households, fiscal measures, and international aid. The "health systems" includes public information campaign, testing policy, contact tracing, emergency investments in health care, investments in COVID-19 vaccines, facial coverings, and vaccination policy.

The information concerning each indicator comprise formal laws, executive orders or other policies issued by regulatory authorities, as well as softer recommendations (guidance and advice). Based on such information, the database team attributes a value $v_{k,t}^i$ which quantifies the intensity of the indicator k at date t .¹ The values can be either ranks on an ordinal scale (for non-financial indicators) or expressed in USD (for financial indicators).

We restrict the set of policies to the "containment and closure" category, to which is added the "facial coverings" indicator. We do so because we wish to scrutinize the link between imposing restrictions on individual behavior and economic preferences

¹For instance, the "Stay-at-home requirements" is encoded as follows: 0 for "No measures", 1 for "recommend not leaving house", 2 for "require not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips", and 3 for "require not leaving house with minimal exceptions (eg allowed to leave once a week, or only one person can leave at a time, etc)".

measured at the individual level within the population. Considering policies that do not have a direct impact on people's daily life would consequently provide a more noisy measure. By the same token, we limit our attention to mandatory policies that likely exert a real influence on the well-being of the individuals. To do so, we consider only policies clearly identified as requirements, and define it as "restrictive policies".² Legally infringing on citizens' liberties is indeed a real political challenge in most countries. On the contrary, recommendations are not binding and may simply correspond to an announcement effect.³

To construct our *Responsiveness* variable for each country, we extract both the date of the first implemented restrictive policy and the number of cases over time from the date of the first reported case within the country to the former date. Our measure of responsiveness is then computed as minus the log number of cases on the day of the first restrictive policy being implemented. We acknowledge that such metric overlooks differences in testing strategies across countries, partly due to an unequal distribution of testing materials around the globe. Alternatively, *Responsiveness* could have been defined with respect to the number of recorded deaths. However, many countries implemented restrictions even before having experienced their first COVID-19 related deaths (Gottlieb et al., 2021). In our sample, only a handful of countries implemented a policy after the first death. The resulting sample size would thus have been ultimately too small to support any meaningful empirical analysis. Furthermore, the number of tests was the main piece of information available to governments when they had to implement their first restrictions, so that it appears empirically sound.

To construct our *Stringency* variable for each country at date t , we apply the

²In practice, the procedure boils down to keeping indicators whose values are above 2. The only exception is the "restrictions on gathering" indicator, for which a value of 1 already implements a restriction on individual behavior.

³The case of the "testing policy" and "contact tracing" indicators, which are excluded in our dataset, deserves some comments. At first sight, they seem to participate in imposing restrictions on individual behavior. However, as emphasized in Hale et al. (2021)'s codebook, the valuation process of these two indicators involve a much bigger part of subjective judgement than the other indicators. This is because the announcement of a testing or a contact tracing policy does not necessarily match having the effective resources to enforce it. Ultimately, whether or not such policies exert a real constraint on individuals is likely to vary substantially over countries. We thus decided to discard these indicators to favor a more homogeneous definition of "restrictive policies".

index-construction procedure from Hale et al. (2021) to our set of restrictive policies. Specifically, based on the values $v_{k,t}^i$, a score is computed for each indicator following two steps. First, the indicators are "flagged" to take into account the geographical scope of the policy. The value of the indicator is reduced by half a rank if the policy targets a specific area instead of being applied nationally.⁴ Second, the indicators are normalized to take into account differences in range across indicators. The flagged value of the indicator is divided by the maximal possible value of the indicator and multiplied by 100. Ultimately, the score for each indicator lies between 0 and 100. Finally, the Stringency Index is computed as the (unweighted) average over all scores of the nine indicators. The computation of the Stringency Index for country i at date t can be summarized as:

$$SI = \frac{1}{K} \sum_{k=1}^K 100 \frac{v_{k,t}^i - 0.5f_{k,t}^i}{N_k^i}$$

with $K = 9$ the number of indicators, $v_{k,t}^i$ the value of indicator k at date t , $f_{k,t}^i$ the flag value for indicator k at date t ($= 1$ if the policy is targeted, 0 otherwise) and N_k^i the maximal value of indicator k .⁵

⁴Not all indicators are flagged in the original dataset because the difference between targeted policy and national policy is not suitable for every indicator. In our dataset, all indicators have a flag, except for international travel bans.

⁵Our Stringency Index is slightly different from the "Stringency Index" proposed by Hale et al. (2021). Basically, our definition is the same as theirs with substituting the public information campaign indicator with the facial coverings indicator. Again the logic for such substitution is to focus on policies imposing an effective restriction on individuals' behavior. While funding public information campaign do not legally force people into adopting a specific behavior, requirements on facial coverings clearly do.

C.1.2 The Global Preferences Survey (GPS)

This section presents the Global Preferences Survey (GPS) and is heavily inspired from Falk and Hermle (2018) since we rely on the same dataset.

C.1.2.1 Content of the survey

The GPS was collected as part of the 2012 Gallup World Poll and contains measures of four fundamental preferences with respect to social and non-social domains: willingness to take financial risks, patience (which captures preferences over the inter-temporal timing of rewards), altruism, and trust.

Except for trust, each preference measure is computed from the answers to two survey items: a quantitative item and a qualitative item. Trust is measured only through a qualitative item. In any case, the selection of the items relies on an ex-ante validation procedure which implies incentivized choices experiments. Such a procedure is detailed in the next sub-section.

The qualitative items elicit participants' subjective evaluation of either how they would act in a certain situation (willingness to act) or what would best describe them as a person (self-assessment). For instance, the qualitative item for risk preferences states: "Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means "completely unwilling to take risks" and a 10 means you are "very willing to take risks". The quantitative items ask participants about hypothetical monetary choice scenarios that mimic standard experimental measures. For instance, the quantitative item for risk preferences provides the participants with a sequence of five interdependent fictitious choices between a fixed and a risky payment (lottery). This allows one to progressively approach the point of indifference between the fixed payment and the lottery, which serves as a revealed preference measure for risk-taking behavior. To account for cross-countries comparability issues, the monetary amounts in such quantitative items are adjusted to correspond to the same share in median income (in the local currency) as the share in the German median income (in the original laboratory experiment). Furthermore, the presence of both qualitative and quantitative items mitigates the risk of culture-specific response behaviors. The table C.1 exposes the wording of the items used in the survey.

Table C.1: Preference measures in the GPS

Preference	Quantitative item	Qualitative item
Willingness to take risks	Sequence of five interdependent questions, following a staircase procedure [†] : “Please imagine the following situation. You can choose between a sure payment of a particular amount of money, or a draw, where you would have an equal chance of getting amount x or getting nothing. We will present to you five different situations. What would you prefer: a draw with a 50 percent chance of receiving amount x , and the same 50 percent chance of receiving nothing, or the amount of y as a sure payment?”	“Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means “completely unwilling to take risks” and a 10 means you are “very willing to take risks”. You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.”
Patience	Sequence of five interdependent questions, following a staircase procedure [†] : “Suppose you were given the choice between receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose. Please assume there is no inflation, i.e., future prices are the same as today’s prices. Please consider the following: Would you rather receive 100 Euro today or x Euro in 12 months?”	“How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?”
Trust	-	”I assume that people have only the best intentions.”
Altruism	“Imagine the following situation: Today you unexpectedly received 1,000 Euro. How much of this amount would you donate to a good cause? (Values between 0 and 1000 are allowed.)”	“How willing are you to give to good causes without expecting anything in return?”

[†] The staircase (or “tree-like”) procedure works as follows: *i*) all subjects are exposed to the same initial price list, *ii*) depending on their choice (A) lottery for risk and later for time against B) safe for risk and today for time) they are exposed to a second price list with a different B) option, and *iii*) they repeat step *ii*) several times.

To obtain the final preference measures from the survey items at the individual level, a two-folded manipulation is executed. First, each of the 7 survey items are z-scored using the entire worldwide sample. Second, these z-scores are averaged using weights obtained as penalized OLS coefficients of observed choices in the experimental validation procedure on the respective survey items. The exact calculation for each preference measure based on the answers to the survey items is the following:

$$\begin{aligned}
 \text{Willingness to take risks} &= 0.4729985 \times \text{Staircase risk} + \\
 &\quad 0.5270015 \times \text{Willingness to take risks} \\
 \text{Patience} &= 0.7115185 \times \text{Staircase patience} + \\
 &\quad 0.2884815 \times \text{Willingness to give up something today} \\
 \text{Trust} &= 1 \times \text{Belief people have best intentions} \\
 \text{Altruism} &= 0.6350048 \times \text{Willingness to give to good causes} + \\
 &\quad 0.3649952 \times \text{Size of donation}
 \end{aligned}$$

C.1.2.2 Item-selection procedure

This sub-section further details the experimental design of the ex-ante validation procedure from Falk et al. (2018) (also see Falk et al., 2016). The objective is to combine the internal validity provided by experimental methods with the external validity provided by large-scale surveys by selecting the items among a large set of survey questions that best predict incentivized choices in the laboratory.

Such a procedure relies on laboratory experiments with a total of 402 student subjects that took place at the Laboratory for Experimental Economics from the University of Bonn in winter 2010/2011. All subjects participated in two 2-hours-long experimental sessions within one week of each other. Each session consisted of two incentivized choice experiments (either on risk and time preferences or on social preferences) and two unrelated surveys (either on social preferences or risk and time preferences). All choice experiments involving social interactions were one-shot and followed a perfect stranger random matching protocol to disentangle social preferences from repeated games motives. Ultimately, each subject made decisions in all four incentivized choice experiments and answered to the corresponding four

surveys. Randomization over the two possibilities for the two sessions for each subject ensured that order effects are accounted for.

In the incentivized choice experiments, the four preferences are measured as follows.

- **Risk preferences** are elicited through two multiple price lists in which subjects chose between a lottery and varying safe options. Resorting to two multiple price lists instead of one aim at reducing the risk of measurement error. Risk-taking is then measured as the average of the two switching rows.
- **Time preferences** are elicited through a similar procedure as risk preferences, except participants had to choose between receiving a payment at the day of the experiment and a larger payment 12 months later.
- **Trust** is elicited as first mover behavior in two investment games (Berg et al., 1995) where the amount sent was either doubled or tripled. Trust is then measured as the average of the two choices.
- **Altruism** is elicited as donation amount in a dictator game with a charitable organization as recipient.

The final measurement of preferences is formalized in the table C.2.

The choice experiments were accompanied by a large set of qualitative and quantitative survey items. The full list of survey items can be found in Falk et al. (2016). The objective of the experimental validation procedure is to include in the GPS the survey items which are the best predictors of incentivized behavior in the choice experiments. Specifically, the selected items for each preference are the best predictors of observed behavior in OLS regressions as evaluated in a stepwise selection approach using the R^2 as fitting criterion over the full set of candidate items.

Table C.2: Preferences elicitation

Preference	Experiment	Measure
Risk Taking	Two multiple price lists in which subjects choose between a lottery and varying safe options	Average of rows in both price lists in which subjects switch from preferring the lottery to the safe option
Patience	Two multiple price lists in which subjects choose between a payment "today" and a larger payment "in 12 months"	Average of rows in two price lists in which subjects switch from preferring the early to the delayed payment
Trust	First mover behavior in two investment games	Average amount sent as a first mover in both investment games
Altruism	First mover behavior in a dictator game with a charitable organization as recipient	Amount of donation

C.1.3 The World Values Survey (WVS)

The World Values Survey is an international survey documenting the evolution over time of individual beliefs, values and motivations across many different societies.⁷ waves have been conducted since 1981. This study relies on the 6th and 7th waves conducted between 2010 and 2014, and between 2016 and 2021, respectively. In addition, the data for the 7th wave is complemented by the data from the European Values Survey, which asks similar questions but is specific to Europe.⁶ We use these surveys as alternative measures of preferences, to distinguish several components of trust (in particular interpersonal versus institutional trust), and to check the sensitivity of the analysis to a change in the countries pool.

- **Risk Preferences.** We construct our measure of risk-taking from the question presenting respondents with a list of items concerning one fictitious person and asking them to indicate for each item "*whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you?*". We consider as willing to take risks a respondent who scores high on the item "*Adventure and taking risks are important to this person; to have an exciting life.*".
- **Times preferences.** We construct our measure of patience from the question presenting respondents with a list of qualities that children can be encouraged to learn at home and asking them to indicate for each quality whether or not they considered it especially important (with a maximum of five qualities). We consider as patient a respondent who selects "*Thrift, saving money and things*".
- **Trust.** We construct three measures of trust from the WVS: two measures of interpersonal trust (narrow and global) and one measure of institutional trust (in the government). The measures of interpersonal trust are based on a question asking respondents to rank how much they trust individual belonging to several groups on 4-points Likert-scale ranging from "not at all" to "a great deal". The groups are "*Your family*", "*Your neighborhood*", "*People*

⁶The wording of the questions that define the preference measures below are the same between the WVS 6th wave and the EVS/WVS.

⁷Such a question is absent from the joint WVS/EVS questionnaire, so that we cannot measure risk preferences in the corresponding dataset.

you know personally”, *People you meet for the first time*”, *People of another religion*”, *People of another nationality*”. The narrow version of interpersonal trust relies only on the response to *People you meet for the first time*”. Such an item indeed appears the closest to the behavioral measure of interpersonal trust in the standard Trust Game (Berg et al., 1995). The global version of interpersonal trust is computed as the sum of responses from all items. The institutional trust measure is obtained from the question asking respondents to rank how much they trust the government on a similar Likert scale.

- **Altruism.** We construct our measure of patience from the question presenting respondents with a list of qualities that children can be encouraged to learn at home and asking them to indicate for each quality whether or not they considered it especially important (with a maximum of five qualities). We consider as altruistic a respondent who selects *Unselfishness*”.

C.1.4 Country-level control variables

In our OLS regression framework, we make use of several country-level variables to distinguish between the role of preferences, country’s characteristics, and the virus’s spread. All these variables are extracted from the Our World In Data database (<https://ourworldindata.org/>).

- **GDP per capita:** Gross Domestic Product at purchasing power parity (constant 2011 international dollars). Available from from the World Bank Development Indicators.
- **Median Age:** median age in the population (projection for 2020). Available from the United Nation Population Division.
- **Population density:** number of people divided by land area, measured in square kilometers. Available from from the World Bank Development Indicators.
- **Worldwide COVID spread:** difference in days between the first (confirmed) worldwide case in our dataset and the first (confirmed) case within the country. Total confirmed cases of COVID-19 (or probable cases, where reported) are available from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.

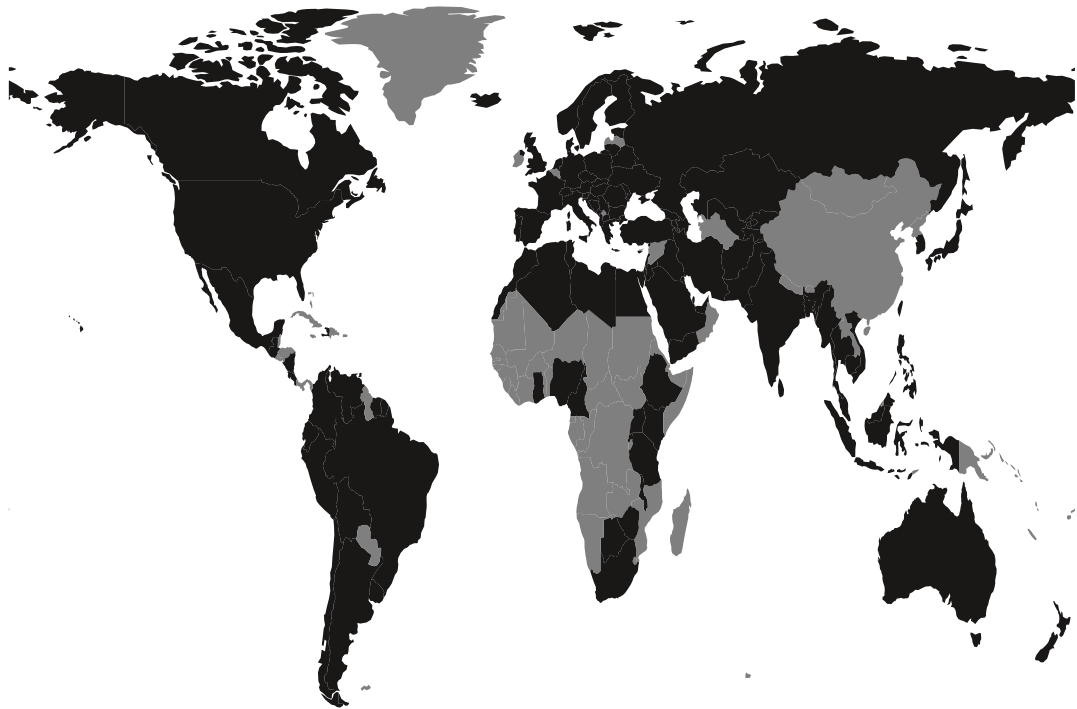
The *GDP per capita*, the *median age* and the *population’s density* aim at controlling for cross-countries differences in the epidemic’s strength. Indeed, an old population will be confronted to a harder epidemic than a younger population and might thus be willing to react sooner and harder to reduce its mortality rate. Similarly, richer countries with efficient health systems will be able to implement lighter restrictions. Finally, regions with higher population density bear the risk that the virus spreads at a higher pace, which is a piece of information that governments are likely to scrutinize.

The *worldwide covid spread* aims at accounting for differences in political reactions over time caused by lately-impacted countries benefiting from first-impacted countries’ experience in dealing with the epidemic (i.e. a ”collective experience effect”). Indeed, the spread of the virus is not random across countries so that the impact of some preferences may be confounded by the contagion order.

C.1.5 List of countries

The figure C.1 and the table C.3 exposes the countries which are included in at least some part of our analysis. Apart from China, which has been excluded because its first recorded case preceeded 2020, our dataset mainly lacks African countries.

Figure C.1: Countries being surveyed at least once



Note: Black identifies countries where at least one survey occurred (GPS, WVS 6th wave or EVS/WVS). Grey identifies countries absent from all our datasets.

Table C.3: Countries and number of respondents per survey

Acronyms	Country	GPS	WVS (6 th wave)	WVS/EVS
AFG	Afghanistan	1000	.	.
ALB	Albania	.	.	1435
AND	Andorra	.	.	1004
ARE	United Arab Emirates	1000	.	.
ARG	Argentina	1000	1030	1003
ARM	Armenia	.	1100	1500
AUS	Australia	1002	1477	1813
AUT	Austria	1001	.	1644
AZE	Azerbaijan	.	1002	1800
BGD	Bangladesh	999	.	1200
BIH	Bosnia and Herzegovina	1004	.	1724
BLR	Belarus	.	1535	1548
BOL	Bolivia	998	.	2067
BRA	Brazil	1003	1486	1762
BWA	Botswana	1000	.	.
CAN	Canada	1001	.	4018*
CHE	Switzerland	1000	.	3174
CHL	Chile	1003	1000	1000
CHN	China	2574	.	.
CMR	Cameroon	1000	.	.
COL	Colombia	1000	1512	1520
CRI	Costa Rica	1000	.	.
CYP	Cyprus	.	1000	1000
CZE	Czech Republic	1005	.	1811
DEU	Germany	997	2046	3698
DNK	Denmark	.	.	3362
DZA	Algeria	1022	1200	.
ECU	Ecuador	.	1202	1200
EGY	Egypt	1020	1523	1200
ESP	Spain	1000	1189	1209
EST	Estonia	1004	1533	1304

ETH	Ethiopia	.	.	1230*
FIN	Finland	1000	.	1199
FRA	France	1001	.	1870
GBR	United Kingdom	1030	.	1788
GEO	Georgia	1000	1202	2194
GHA	Ghana	1000	1552	.
GRC	Greece	1000	.	1200
GTM	Guatemala	1000	.	1203*
HKG	Hong Kong	.	1000	2075
HRV	Croatia	992	.	1487
HTI	Haiti	504	1996	.
HUN	Hungary	1004	.	1514
IDN	Indonesia	1000	.	3200
IND	India	2539	4078	.
IRN	Iran	2507	.	1499*
IRQ	Iraq	1000	1200	1200
ISL	Iceland	.	.	1624
ITA	Italy	1004	.	2277
JOR	Jordan	1000	1200	1203
JPN	Japan	1000	2443	1353
KAZ	Kazakhstan	999	1500	1276
KEN	Kenya	1000	.	.
KHM	Cambodia	1000	.	.
KOR	South Korea	1000	1200	1245
KWT	Kuwait	.	1303	.
KGZ	Kyrgyzstan	.	1500	1200*
LBN	Lebanon	.	1200	1200
LBY	Libya	.	2131	.
LKA	Sri Lanka	1000	.	.
LTU	Lithuania	999	.	1448
MAR	Morocco	1000	1200	.
MDA	Moldova	1000	.	.
MEX	Mexico	1000	2000	1739
MKD	North Macedonia	.	.	1117
MMR	Myanmar	.	.	1200*

APPENDIX C

MNE	Montenegro	.	.	1003
MWI	Malawi	1000	.	.
MYS	Malaysia	.	1300	1313
NGA	Nigeria	1000	1759	1237
NIC	Nicaragua	1000	.	1200*
NLD	Netherlands	1000	1902	2404
NOR	Norway	.	.	1122
NZL	New Zealand	.	841	1057*
PAK	Pakistan	1004	1200	1995
PER	Peru	1000	1210	1400
PHL	Philippines	1000	1200	1200
POL	Poland	999	966	1352
PRT	Portugal	998	.	1215*
PSE	Palestine	.	1000	.
QAT	Qatar	.	1060	.
ROU	Romania	994	1503	2870
RUS	Russian Federation	1498	2500	3635
RWA	Rwanda	1000	1527	.
SAU	Saudi Arabia	1035	.	.
SGP	Singapore	.	1972	2012*
SVK	Slovakia	.	.	1432
SVN	Slovenia	.	1069	1075
SRB	Serbia	1023	.	2545
SUR	Suriname	504	.	.
SWE	Sweden	1000	1206	1194
THA	Thailand	1000	1200	1500
TJK	Tajikistan	.	.	1200*
TTO	Trinidad and Tobago	.	999	.
TUN	Tunisia	.	1205	1208
TUR	Turkey	1000	1605	2415
TWN	Taiwan	.	1238	1223
TZA	Tanzania	1000	.	.
UGA	Uganda	1000	.	.
UKR	Ukraine	1000	1500	2901*
URY	Uruguay	.	1000	.

USA	United States	1072	2232	2596	
UZB	Uzbekistan	.	1500	.	
VEN	Venezuela	999	.	.	
VNM	Vietnam	1000	.	1200*	
YEM	Yemen	.	1000	.	
ZAF	South Africa	1000	3531	.	
ZWE	Zimbabwe	1000	1500	1215*	
Number of countries ⁷		109	75	59	78
Number of respondents			79,338	86,365	128,256

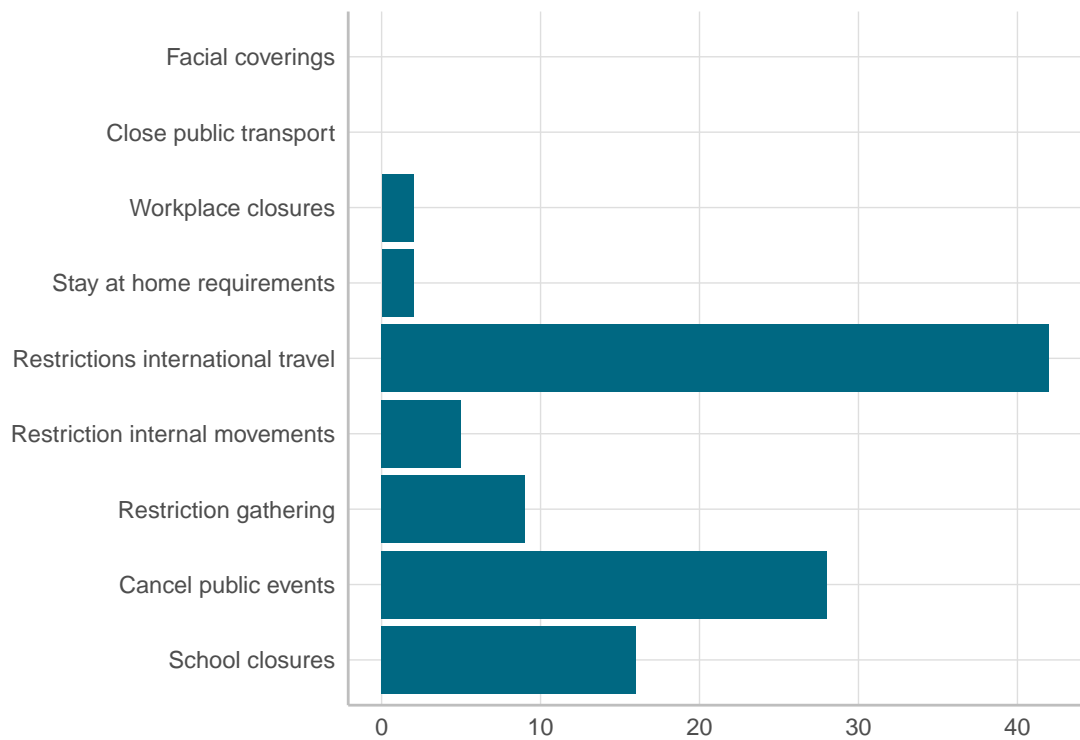
⁷Some countries from the original surveys are missing due to the unavailability of the data in the OxcGRT dataset. All datasets exclude China. The WVS/EVS dataset also excludes Puerto Rico and Macau SAR. Countries tagged with * have been surveyed after 2020 in the EVS/WVS.

C.2 Descriptive statistics

C.2.1 Statistics on dependant variables

The figure C.2 depicts the proportion of each of the nine restrictive policies among the policies which were implemented first by governments in reaction to the COVID-19 epidemic. International travel bans are the dominant reaction by far, but we also observe some heterogeneity in policy-making, as a primary indication that not all countries reacted in the same way.

Figure C.2: First restrictive policy being implemented



In the figures C.3, C.4 and C.5, we map the value of our dependant variables (*Responsiveness* and *Stringency*) over different datasets and measurements (for *Responsiveness*) or time (for *Stringency*).

Concerning *Responsiveness*, we consider two alternative definitions as robustness checks. As a reminder, our main definition for *Responsiveness* is: $-\log(\text{Number of recorded cases on the day of the first restrictive policy being implemented})$ (definition R_1)).

Definition R_2 : – Number of days between the first recorded case and the day when the first restrictive policy is implemented. The more days pass before the first restrictive policy, the less responsive the government is.

Definition R_2 corresponds to the most intuitive definition of governments' *Responsiveness* as the time spent before intervening. However, such a definition does not take into account that the virus spreads at different paces across countries (see e.g. Alimohamadi et al., 2020), which results in different needs for early intervention. On the contrary, both definitions R_1 and R_3 account for the local dynamics of the epidemic.

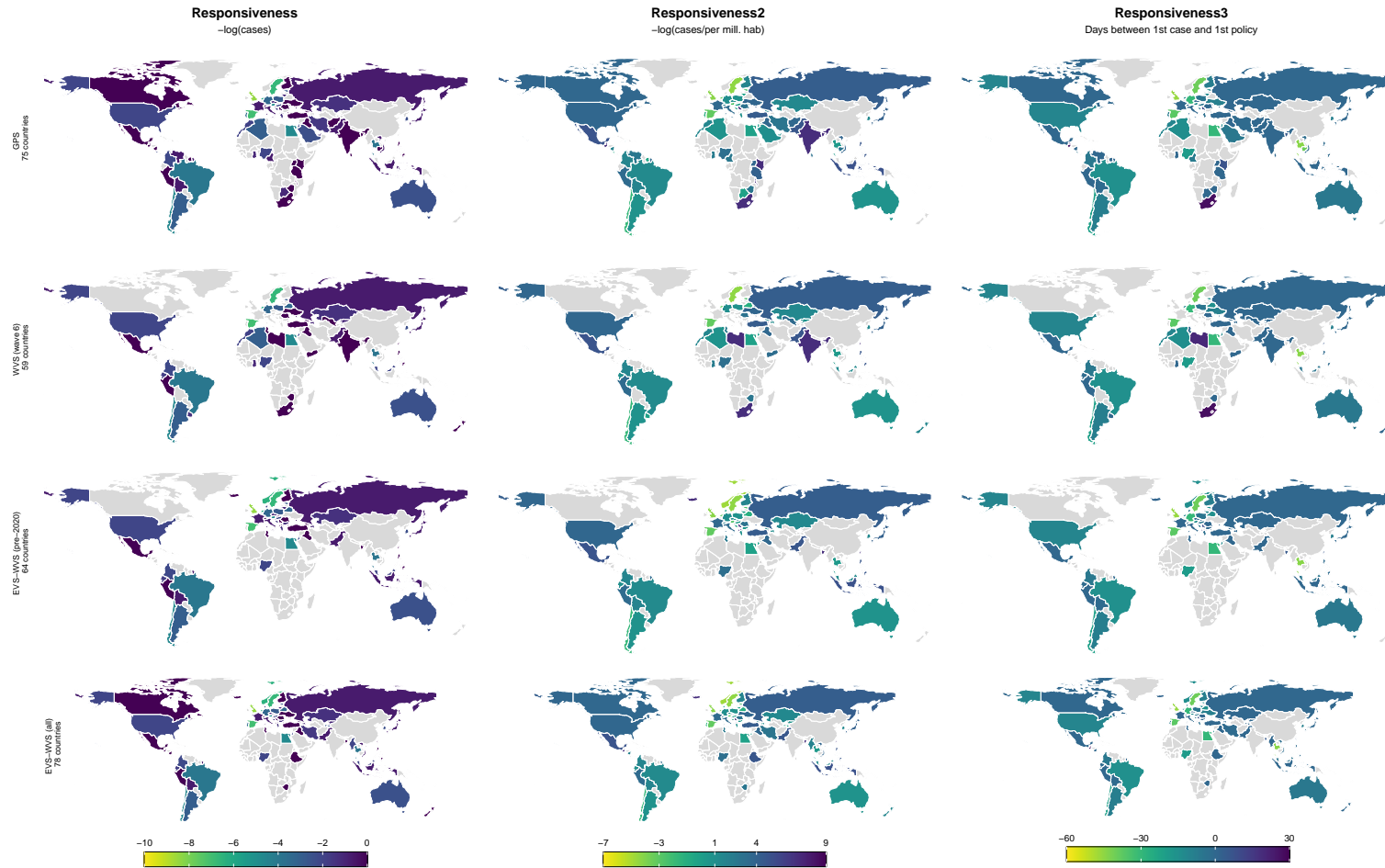
Definition R_3 : $-\log(\text{Number of recorded cases per million inhabitants on the day of the first restrictive policy being implemented})$. The higher the share of recorded cases in the total population (per millions), the less responsive the government is.

Definition R_3 is based on the share of recorded cases within the population instead of considering the absolute number of recorded cases (as in R_1). It thus accounts for how important the epidemic is with respect to the entire population. However, the relative importance of the epidemic (through so-called "incidence rates") was not heavily scrutinized by the governments or the media at the beginning of the epidemic as compared to the absolute number of cases.

Concerning *Stringency*, we map the index for every three months starting on the day of the first restrictive policy being implemented.

Overall, we observe substantial heterogeneity throughout the world in terms of both *Responsiveness* and *Stringency*, no matter the dataset under scrutiny. In particular, although all countries appear to have converged towards a high-level response, we still find different stringency indices at the same moment in terms of the local dynamics of the epidemic, and different patterns over time during the waxing and waning phase. Besides, our different measures for *Responsiveness* yield globally consistent results.

Figure C.3: *Responsiveness* over data-sets and measurements



336

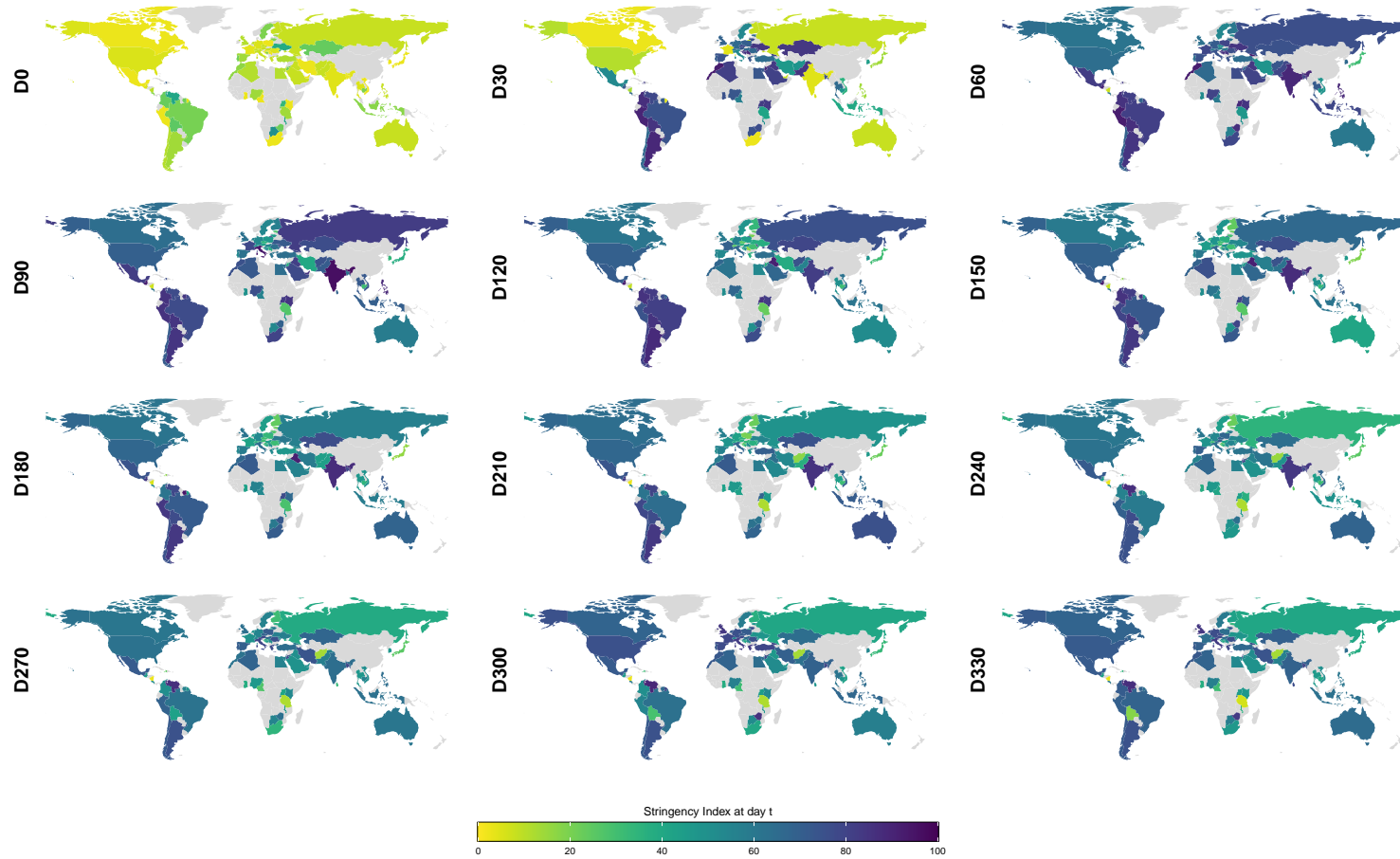
APPENDIX C

Note: *Responsiveness* on the left is computed as minus the log number of recorded cases on the day when the first restrictive policy is implemented. For instance, a value of -4 corresponds to approximately 55 recorded cases.

Responsiveness in the center is computed as minus the log number of recorded cases per millions inhabitants on the day when the first restrictive policy is implemented. For instance, a value of 1 corresponds to approximately 3 recorded cases per millions inhabitants.

Responsiveness on the right is computed as minus the number of days between the first recorded case and the day when the first restrictive policy is implemented. For instance, a value of 0 means that the first restrictive policy was implemented on the day the country recorded its first case.

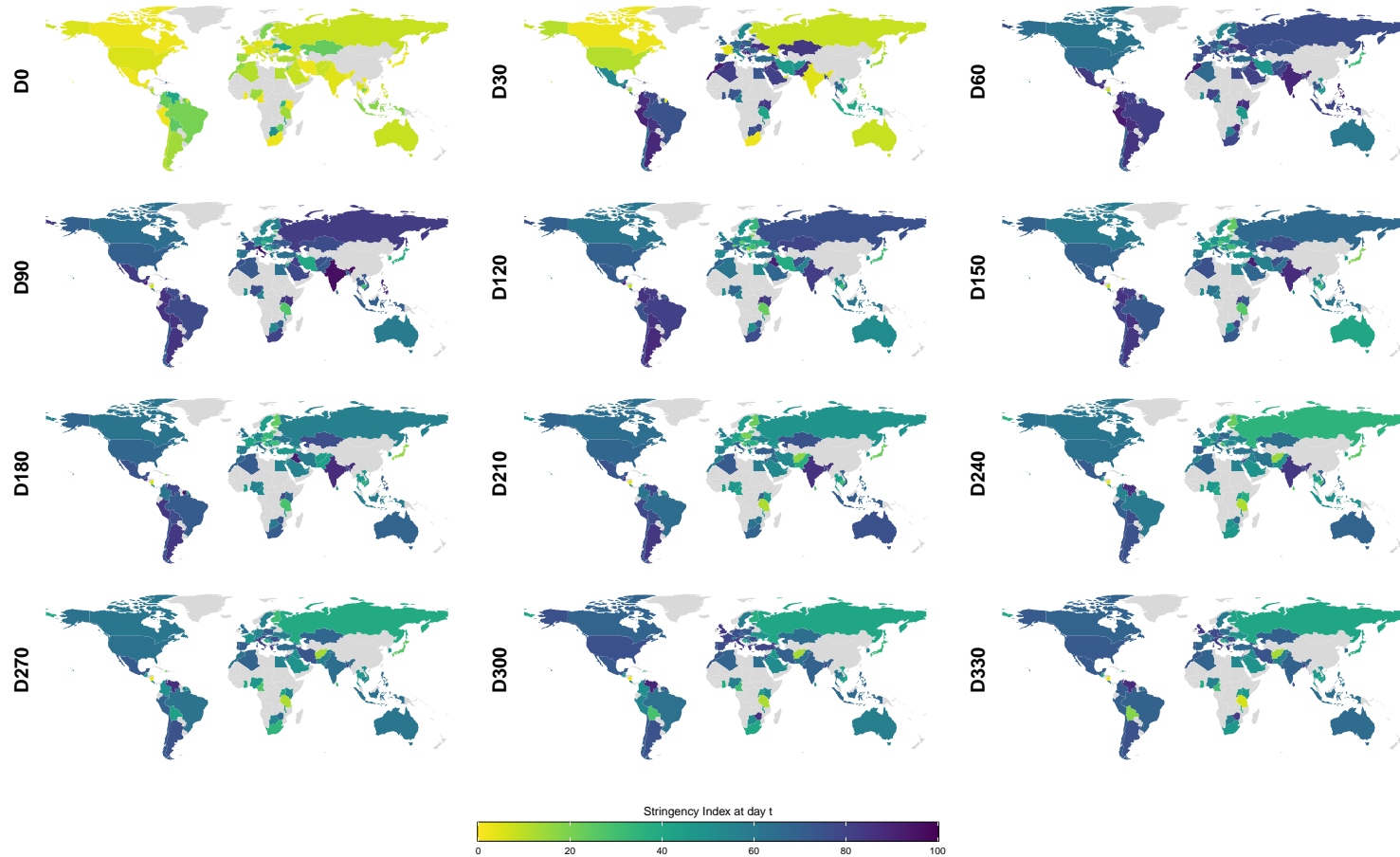
Figure C.4: *Stringency* at different time spans (GPS)



Note: *Stringency* is an index ranging from 0 (slightly-coercive response) to 100 (highly-coercive response). It is measured every day following the day where the first restrictive policy is implemented (e.g., D_{30} is D_0 plus 30, hence one month after the first restrictive policy being implemented).

Figure C.5: *Stringency* at different time spans (WVS)

338



APPENDIX C

Note: *Stringency* is an index ranging from 0 (slightly-coercive response) to 100 (highly-coercive response). It is measured every day following the day where the first restrictive policy is implemented (e.g., D_{30} is D_0 plus 30, hence one month after the first restrictive policy being implemented).

C.2.2 Relationships between independent variables

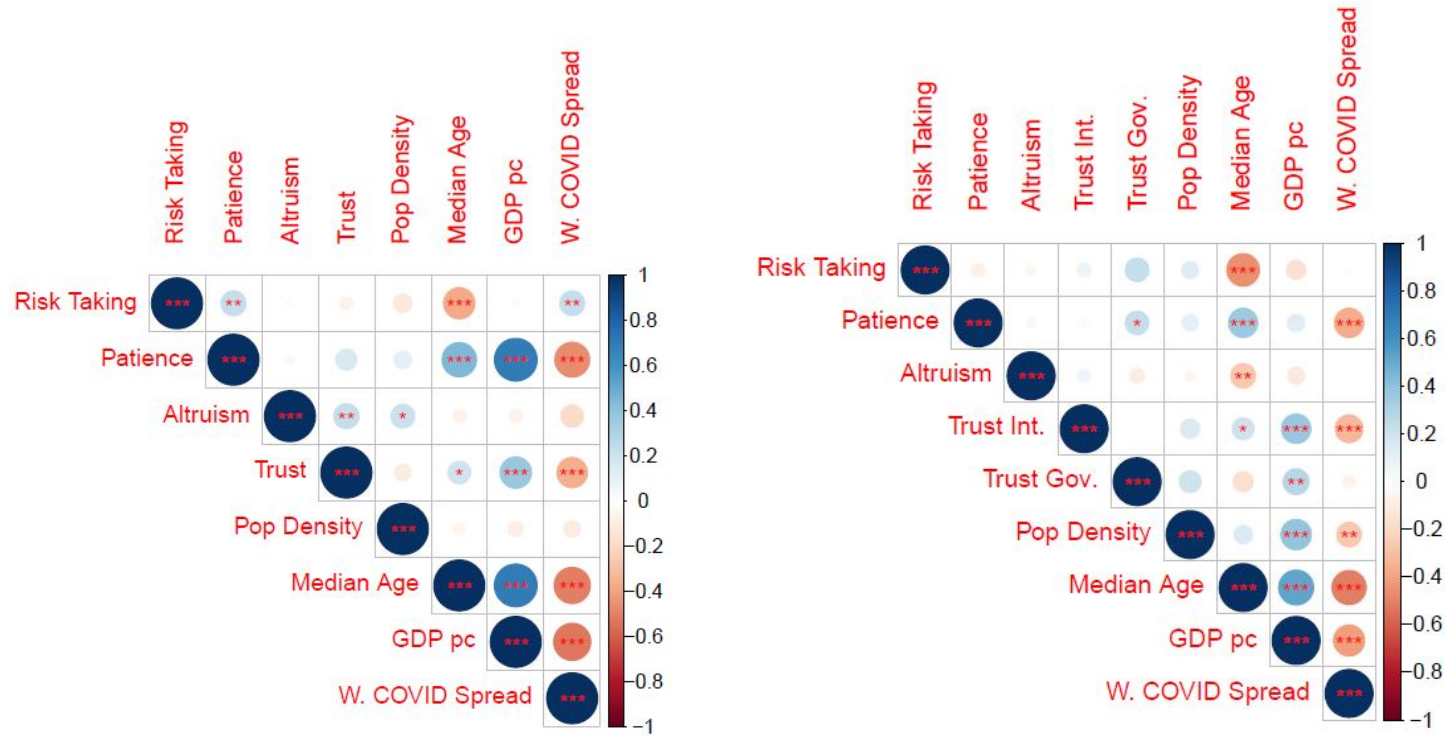
In this section, we correlate our independent variables within (figure C.6) and between (figure C.7) data sources.

In the figure C.6, we document the relationships between all our independent variables both within the GPS (left panel) and within the WVS (right panel). We observe that populations' preferences are related to their socio-demographic characteristics and to the temporal distance with the first worldwide case. In particular, older populations tend to be more risk-averse and more patient than younger ones, while wealthier populations tend to be more patient and more trustful than poor populations. Furthermore, patient, risk-averse and trustful populations have been affected first by the virus. Globally, such results are consistent between the GPS and the WVS. These correlations confirm the need to take into account in our regression framework the non-random spread of the virus, which quickly affected the western countries which tend to be older and wealthier than the other countries.

In the figure C.7, we document the relationships between our preference measures across datasets: GPS v. WVS 6th wave (left panel) and GPS v. EVS/WVS (right panel). Globally, we observe that our different datasets measure similar preferences: although not always significant, the between-datasets correlations for risk-taking, patience and interpersonal trust are positive. Interestingly, we observe a negative correlation for altruism in both comparisons. Such a finding may indicate that the surveys ultimately quantify different notions of "selfishness". It is indeed possible that donating to a good cause (GPS measure) does not encompass all dimensions of altruistic behavior. At the same time, it is also possible that the very mention of "selfishness" in a declarative survey item (as it is the case in the WVS) provides a free lunch opportunity for selfish individuals to signal themselves as altruistic, which is often considered a morally superior personality trait. Finally, we also observe that some of our preference measures are correlated with each other (e.g., altruism and trust), hence the need to display results both with a control of other preference measures and without any control as a robustness check.

Figure C.6: Correlations within data sources

340

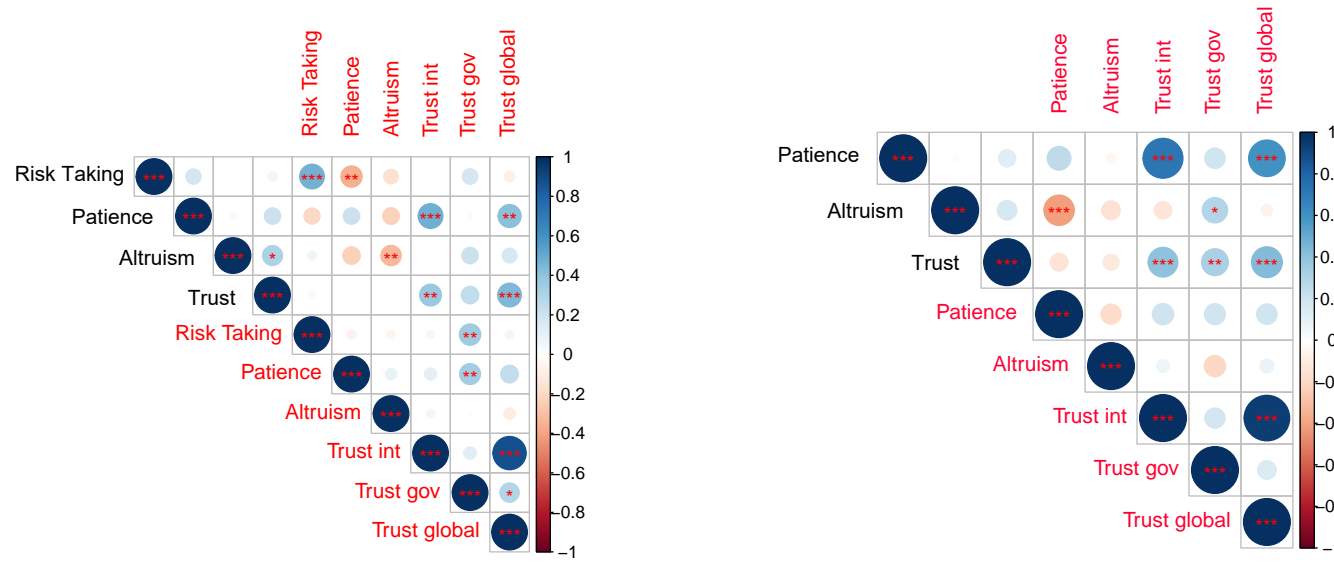


(a) Correlations between variables (GPS)

(b) Correlations between variables (WVS)

Note: The figure C.6a (resp. C.6b) presents correlations between the independent variables we focus on in the GPS (resp. WVS) database.

Figure C.7: Correlation between data source (GPS, WVS 6th wave and WVS/EVS)



(a) GPS vs. WVS 6th wave correlations

(b) GPS v. WVS/EVS correlations

Note: The figure C.7a presents correlations between our preference measures in the GPS (black) and in the WVS 6th (red). The sample consists of 37 countries which take part in both surveys. The figure C.7b presents correlations between our preference measures in the GPS (black) and in the WVS/EVS (red). The sample consists of 42 countries which take part in both surveys.

C.3 Model specification

C.3.1 OLS regression on *Responsiveness*

$$Y_i = \alpha + \beta \times \text{Pref}_i + \gamma \times \text{Other pref}_i + \theta \times \text{Controls} + \epsilon_i$$

where:

- Y_i is the outcome variable for country i , which is *Responsiveness* ($-\log$ of number of cases on the day of the first implemented restriction)
- Pref is the economic preference under scrutiny and β is the corresponding regression coefficient.
- Other pref are the other economic preferences and γ is the corresponding vector of regression coefficients.
- X_i are the other variables controlled for (GDP/capita, median age, populations' density and worldwide covid spread) and θ is the corresponding vector of regression coefficients.
- ϵ_i is the error term.

C.3.2 Fixed effect panel regression on *Stringency*

$$Y_{it} = \sum_{t=1}^{t=365} \beta_t \times (\text{Pref}_i \times D_t) + \sum_{t=1}^{t=365} \gamma_t \times (\text{Other pref}_i \times D_t) + \theta \times \log(\text{Cases}_{it}) + \alpha_i + D_t + \epsilon_{it}$$

where:

- Y_{it} is the outcome variable for country i at date t , which is the *Stringency* (Stringency Index)
- D_t is a dummy for date t
- Pref is the economic preference under scrutiny and $\beta_1 \dots \beta_{365}$ are the associated regression coefficients over all dates t .
- Other pref are the other economic preferences and $\gamma_1 \dots \gamma_{365}$ are the associated regression coefficients over all dates t .
- α_i denotes countries' fixed effect.
- ϵ_{it} is the error term.

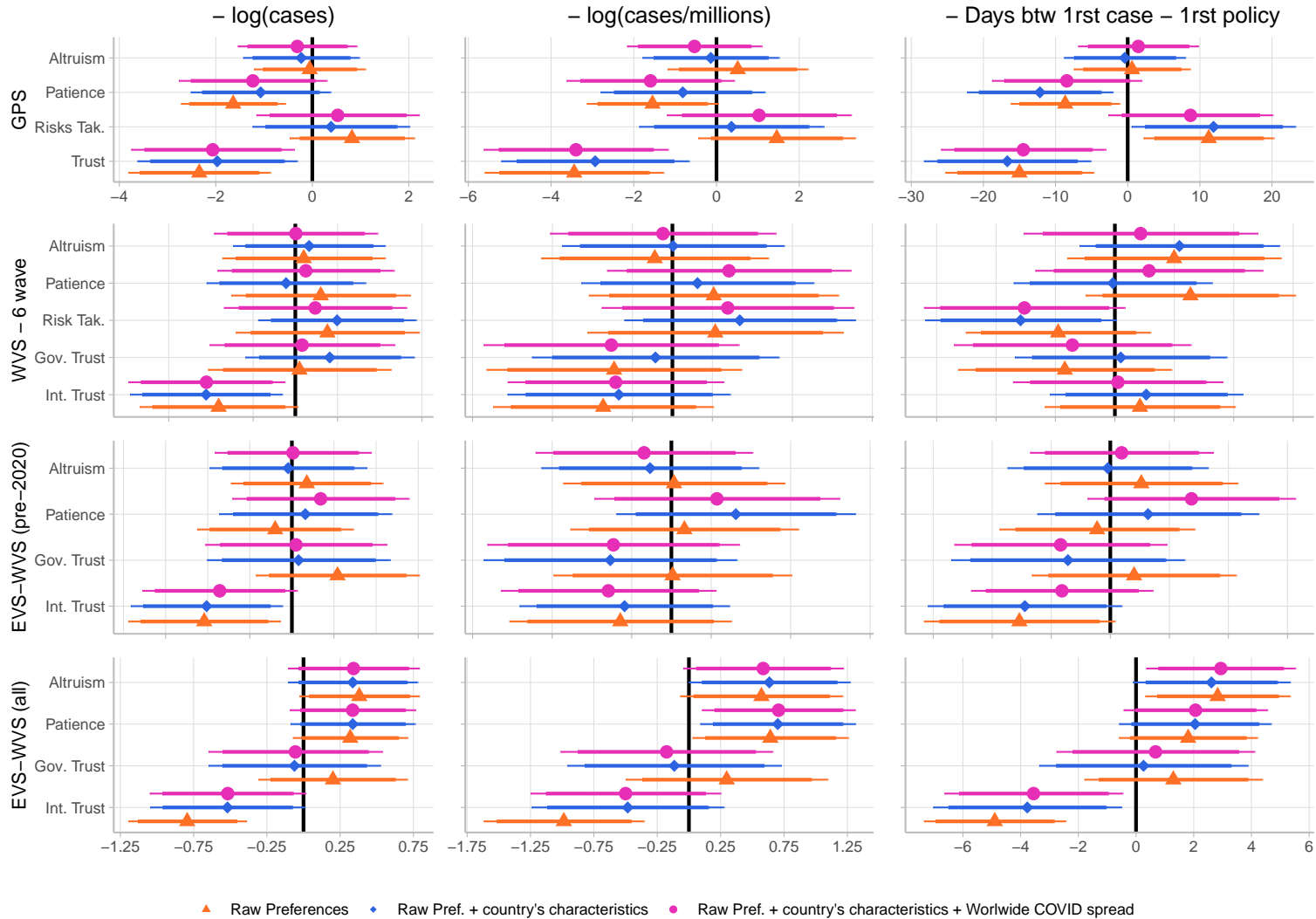
C.4 Robustness checks

C.4.1 *Responsiveness: alternative datasets and definitions*

We replicate our analysis on governments' *Responsiveness* considering our two alternative definitions R_2 and R_3 . The results obtained when varying the definition of *Responsiveness* on all four datasets (GPS, WVS 6th wave, full EVS/WVS and restricted EVS/WVS on countries surveyed before 2020) are exposed in the figures C.8 and C.9.

In the figure C.8, we consider the narrow definition of trust (self-assessment on whether strangers are trustworthy) in both the WVS 6th wave and the EVS/WVS. In the figure C.9, we alternatively consider the global definition of trust (self-assessment on whether various groups of people are trustworthy, and summing the scores for all groups).

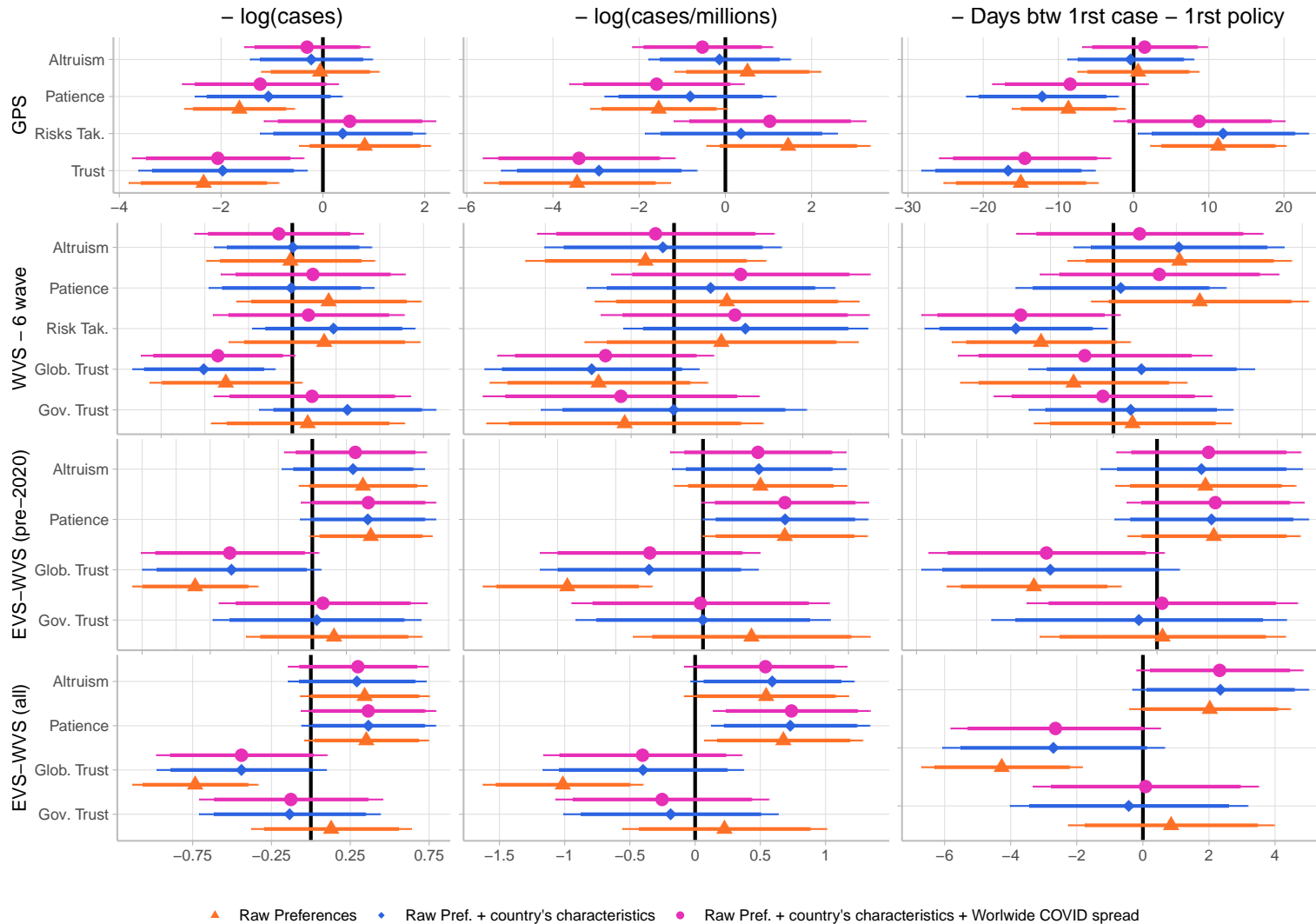
Figure C.8: Alternative specifications of *Responsiveness* - Narrow Trust



Note: Each dot corresponds to the estimated coefficient associated with a given preference measure in OLS regressions of governments' *Responsiveness* on nationally-aggregated preferences. The bars correspond to 90% (shorter) and 95% (larger) confidence intervals, computed with robust standard errors. If one bar overlaps 0, the corresponding coefficient is not statistically significant at the given confidence level. The regressions include either only the preference measures ("raw preferences") or add some controls. These controls are *population density*, *median age*, $\log(GDP)_{pc}$ ("country's characteristics") and *worldwide COVID spread*. The figure organizes as follow. Each column represents a given measure of Responsiveness : column 1 is the $-\log(\text{number of cases})$, 2 is the $-\log(\text{number of cases}/\text{per millions inhabitants})$, and 3 the number of days between the first case and the first required policy. Each row gathers estimates for each measure of *Responsiveness* for a given data-set. See Appendices C.1.2 and C.1.3 for details on variables' definition.

Figure C.9: Alternative specification of *Responsiveness* - Global Trust

346



APPENDIX C

Note: Each dot corresponds to the estimated coefficient associated with a given preference measure in OLS regressions of governments' *Responsiveness* on nationally-aggregated preferences. The bars correspond to 90% (shorter) and 95% (larger) confidence intervals, computed with robust standard errors. If one bar overlaps 0, the corresponding coefficient is not statistically significant at the given confidence level. The regressions include either only the preference measures ("raw preferences") or add some controls. These controls are *population density*, *median age*, $\log(GDP)_{pc}$ ("country's characteristics") and *worldwide COVID spread*. The figure organizes as follow. Each column represents a given measure of Responsiveness : column 1 is the $-\log(\text{number of cases})$, 2 is the $-\log(\text{number of cases/per millions inhabitants})$, and 3 the number of days between the first case and the first required policy. Each row gathers estimates for each measure of *Responsiveness* for a given data-set. See Appendices C.1.2 and C.1.3 for details on variables' definition.

C.4.2 *Responsiveness*: Full regression tables

The following tables expose our results on *Responsiveness* in a series of regression models which vary the number of dependant variables incrementally. Our results on preferences remain qualitatively similar as the one exposed in the main text throughout all specifications.

The results with the GPS data are exposed in table C.4. The first four columns regress preferences on *Responsiveness* one at the time and the fifth column includes all preferences in the model. The last three columns add our control variables incrementally.

The results with the WVS data are exposed in tables C.5 (narrow trust) and C.6 (global trust). Along with the WVS version of the four preferences present in the GPS, we add a measure of trust in the government. The first five columns thus regress preferences on *Responsiveness* one at the time. The remaining columns implement a similar model as in columns (4) to (8) from table C.4, with sandwiched models in the middle that simply add trust in the government.

Table C.4: Determinants of *Responsiveness* (GPS data)

	<i>Dependent variable:</i>							
	<i>Responsiveness</i> (- log(number of cases when first restrictive policy))							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Willing. to take risks	0.482 (0.526)				0.821* (0.468)	0.551 (0.486)	0.389 (0.641)	0.526 (0.683)
Patience		-1.766** (0.742)			-1.640** (0.688)	-1.048 (0.838)	-1.071 (0.899)	-1.232 (0.953)
Trust			-2.785*** (0.717)		-2.343*** (0.606)	-1.875*** (0.707)	-1.969*** (0.755)	-2.065*** (0.743)
Altruism				-0.452 (0.505)	-0.055 (0.434)	-0.178 (0.419)	-0.230 (0.455)	-0.311 (0.471)
GDP_{pc} in log						-0.387** (0.191)	-0.196 (0.340)	-0.228 (0.350)
Median age							-0.023 (0.037)	-0.025 (0.037)
Pop. density							0.0005 (0.001)	0.0004 (0.001)
Worldwide COVID spread								-0.009 (0.013)
Constant	-1.468*** (0.218)	-1.477*** (0.209)	-1.545*** (0.215)	-1.482*** (0.220)	-1.559*** (0.209)	2.144 (1.841)	1.032 (2.543)	1.683 (2.745)
Observations	75	75	75	75	75	75	75	75
R ²	0.006	0.120	0.162	0.007	0.259	0.282	0.287	0.292
Adjusted R ²	-0.008	0.108	0.150	-0.007	0.216	0.230	0.213	0.206

Note: The dependant variable is minus the log number of cases on the day of the first implemented policy (*Responsiveness*). Estimates are obtained through Ordinary Least Squares (OLS). The four preferences come from the Global Preferences Survey. GDP per capita, median age and population density comes from the OxCGRT. The worldwide covid spread corresponds to the number of days between the first worldwide case and the first case in the country. Significance: *p<0.1, **p<0.05, ***p<0.01. Robust standard errors in parentheses.

Table C.5: Determinants of *Responsiveness* - narrow trust (WVS data)

	Dependent variable:												
	<i>Responsiveness</i> (- log(number of cases when first restrictive policy))												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Willing. to take risks	0.271 (0.201)					0.296 (0.218)	0.247 (0.219)	0.342 (0.235)	0.282 (0.233)	0.118 (0.273)	0.118 (0.275)	0.190 (0.272)	0.190 (0.273)
Patience		-0.072 (0.200)				-0.003 (0.205)	-0.055 (0.206)	0.007 (0.193)	-0.040 (0.199)	0.077 (0.203)	0.062 (0.220)	0.160 (0.225)	0.151 (0.240)
Trust Int.			-0.520* (0.285)			-0.530* (0.280)	-0.529* (0.286)	-0.540** (0.258)	-0.530** (0.267)	-0.526** (0.254)	-0.527** (0.256)	-0.454* (0.252)	-0.455* (0.253)
Altruism				0.048 (0.185)		0.059 (0.206)	0.081 (0.211)	0.078 (0.189)	0.091 (0.196)	-0.005 (0.205)	0.003 (0.219)	0.045 (0.209)	0.050 (0.219)
Trust Gov.					0.237 (0.206)		0.204 (0.208)		0.212 (0.227)		0.040 (0.226)		0.025 (0.208)
GDP_{pc} in log								-0.366 (0.274)	-0.395 (0.254)	-0.200 (0.395)	-0.221 (0.378)	-0.070 (0.389)	-0.083 (0.384)
Median age										-0.057 (0.061)	-0.053 (0.062)	-0.042 (0.056)	-0.040 (0.058)
Pop. density										0.0004*** (0.0001)	0.0004*** (0.0001)	0.0004** (0.0002)	0.0004** (0.0002)
Worldwide COVID spread												0.023 (0.015)	0.023 (0.015)
Constant	-1.494*** (0.232)	-1.491*** (0.233)	-1.516*** (0.227)	-1.490*** (0.233)	-1.481*** (0.229)	-1.522*** (0.230)	-1.513*** (0.229)	1.906 (2.588)	2.193 (2.396)	2.094 (2.548)	2.176 (2.438)	-0.456 (2.956)	-0.400 (2.902)
Observations	59	59	58	59	59	58	58	56	56	56	56	56	56
R ²	0.023	0.002	0.086	0.001	0.016	0.116	0.126	0.206	0.218	0.270	0.271	0.303	0.303
Adjusted R ²	0.006	-0.016	0.070	-0.017	-0.001	0.049	0.042	0.127	0.122	0.164	0.146	0.185	0.167

Note: The dependant variable is minus the log number of cases on the day of the first implemented policy (*Responsiveness*). The sample consists of all countries being surveyed in the 6th wav of the WVS. Estimates are obtained through Ordinary Least Squares (OLS). The four preferences come from the World Values Survey. The trust variable corresponds to the narrow definition of trust in the WVS (trust in strangers). GDP per capita, median age and population density come from the OxCGRT. The worldwide covid spread corresponds to the number of days between the first worldwide case in our dataset and the first case in the country. Significance: *p<0.1, **p<0.05, ***p<0.01. Robust standard errors in parentheses.

Table C.6: Determinants of *Responsiveness* - global trust (WVS data)

	Dependent variable:												
	<i>Responsiveness</i> (- log(number of cases when first restrictive policy))												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Willing. to take risks	0.271 (0.201)					0.307 (0.209)	0.235 (0.208)	0.353 (0.230)	0.276 (0.226)	0.089 (0.266)	0.092 (0.269)	0.180 (0.265)	0.182 (0.268)
Patience		-0.072 (0.200)				0.065 (0.215)	-0.006 (0.217)	0.057 (0.208)	-0.009 (0.213)	0.159 (0.217)	0.117 (0.236)	0.241 (0.235)	0.208 (0.252)
Trust global			-0.505* (0.270)			-0.530* (0.283)	-0.579** (0.291)	-0.479* (0.288)	-0.518* (0.302)	-0.470* (0.280)	-0.487* (0.285)	-0.422 (0.267)	-0.435 (0.273)
Altruism				0.048 (0.185)		-0.025 (0.209)	0.003 (0.217)	0.006 (0.197)	0.026 (0.207)	-0.100 (0.210)	-0.077 (0.225)	-0.028 (0.213)	-0.011 (0.225)
Trust Gov.					0.237 (0.206)		0.315 (0.215)		0.306 (0.229)		0.113 (0.233)		0.088 (0.212)
GDP_{pc} in log								-0.304 (0.278)	-0.322 (0.262)	-0.043 (0.412)	-0.096 (0.399)	0.090 (0.400)	0.047 (0.398)
Median age										-0.075 (0.061)	-0.064 (0.063)	-0.055 (0.056)	-0.047 (0.059)
Pop. density										0.0004*** (0.0001)	0.0003*** (0.0001)	0.0004*** (0.0001)	0.0004*** (0.0001)
Worldwide COVID spread												0.026* (0.014)	0.026* (0.014)
Constant	-1.494*** (0.232)	-1.491*** (0.233)	-1.514*** (0.227)	-1.490*** (0.233)	-1.481*** (0.229)	-1.520*** (0.230)	-1.505*** (0.226)	1.311 (2.630)	1.505 (2.488)	1.201 (2.689)	1.364 (2.596)	-1.657 (3.092)	-1.502 (3.026)
Observations	59	59	58	59	59	58	58	56	56	56	56	56	56
R ²	0.023	0.002	0.081	0.001	0.016	0.111	0.136	0.178	0.202	0.248	0.251	0.292	0.293
Adjusted R ²	0.006	-0.016	0.065	-0.017	-0.001	0.044	0.053	0.096	0.104	0.139	0.123	0.171	0.155

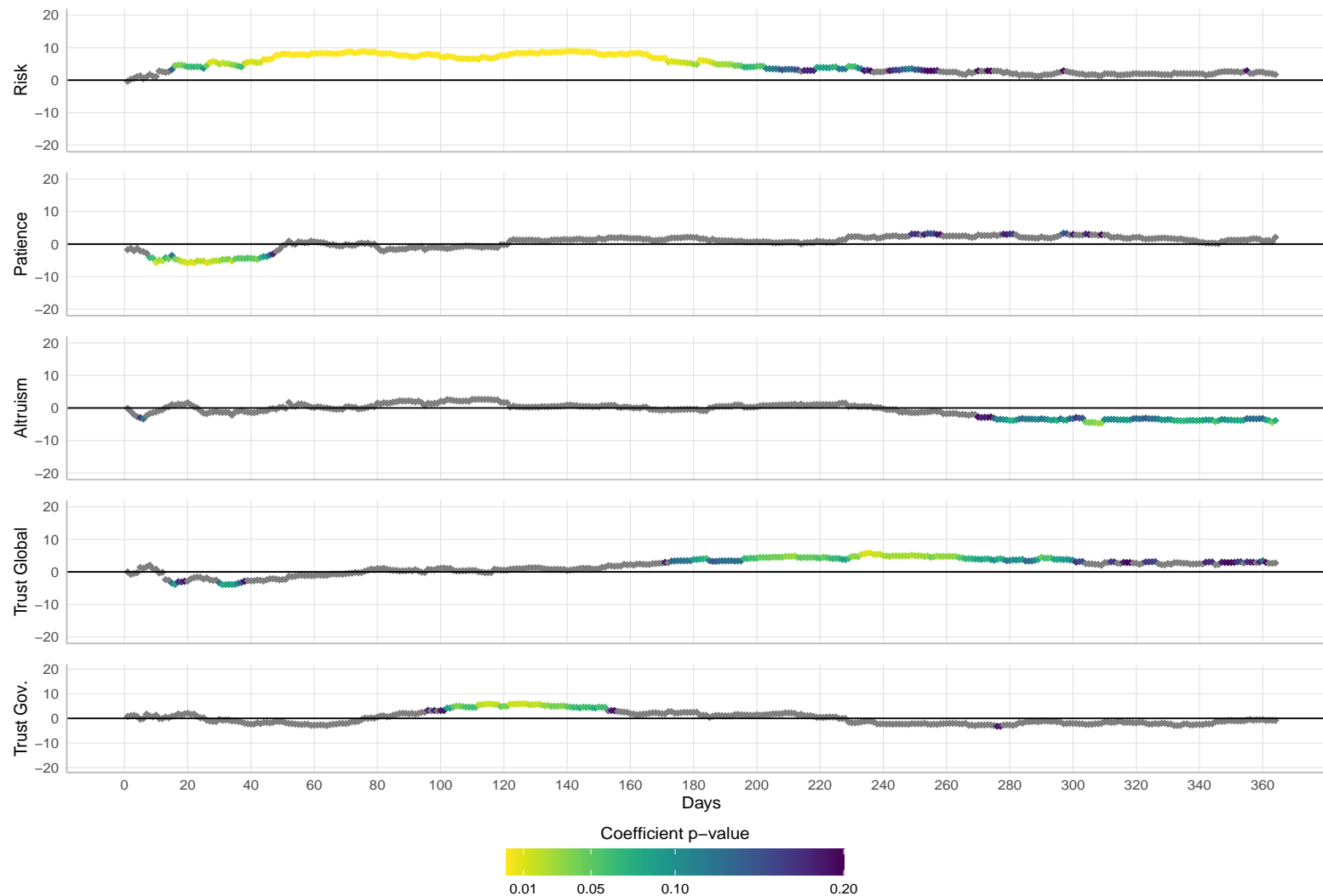
Note: The dependant variable is minus the log number of cases on the day of the first implemented policy (*Responsiveness*). The sample consists of all countries being surveyed in the 6th wave of the WVS. Estimates are obtained through Ordinary Least Squares (OLS). The four preferences come from the World Values Survey. The trust variable corresponds to the global definition of trust in the WVS (sum of scores from questions on trust). GDP per capita, median age and population density come from the OxCGRT. The worldwide covid spread corresponds to the number of days between the first worldwide case in our dataset and the first case in the country. Significance: *p<0.1, **p<0.05, ***p<0.01. Robust standard errors in parentheses.

C.4.3 *Stringency*: Additional results using WVS data

In this section, we complement our analysis on *Stringency* using the WVS 6th wave by considering the global definition of trust (self-assessment on whether various groups of people are trustworthy, and summing the scores for all groups) in the figure C.10, or by running the analyses on the EVS/WVS database in the figure C.11. Overall we observe consistent patterns with our main results.

Figure C.10: *Stringency* and preferences - WVS 6th wave (Global Trust)

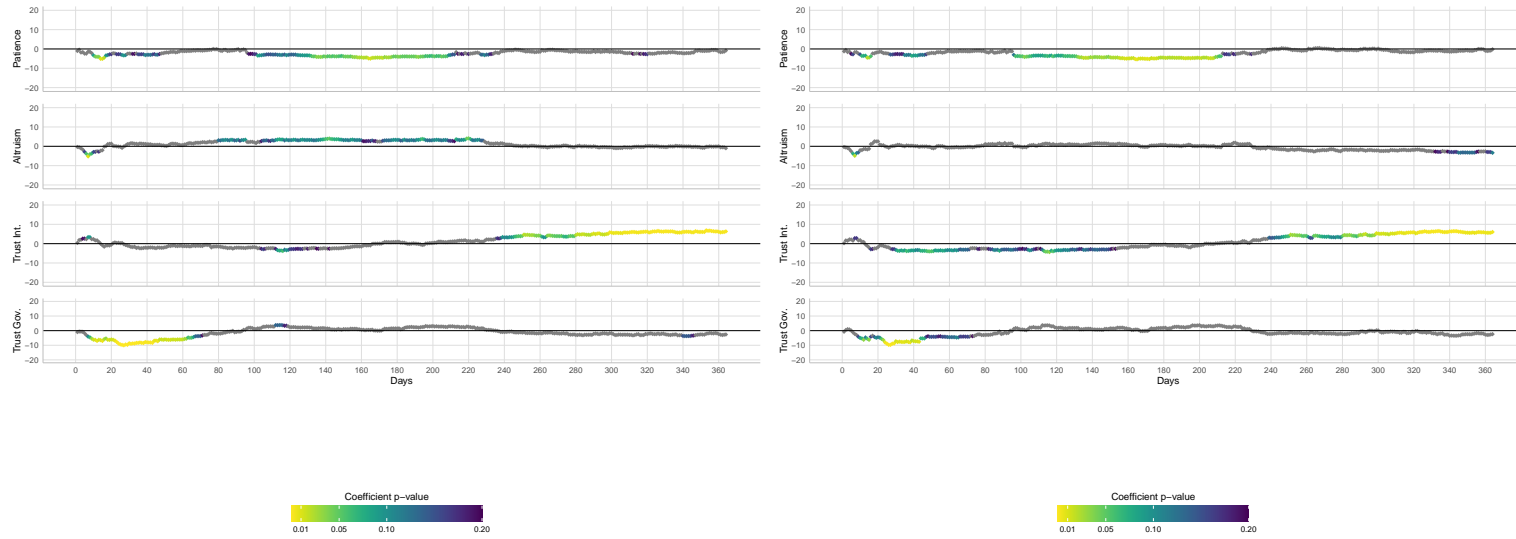
352



APPENDIX C

Note: The figure illustrates 364 coefficients from a panel regression of economic preferences on the Stringency Index at day D_t ($t \in [1, 365]$) along with several control variables using the WVS 6th wave dataset. Each regression includes all preference variables (*e.g.* risk-taking, patience, altruism or trust) interacted with the day dummies D_t as well as the (log) number of cases at date t , the day dummies and country-level fixed effects. The graph for each preference summarizes the information obtained from the regressions about the corresponding preference. The dot for date t conveys two pieces of information: the position indicates both the sign and magnitude of the coefficient associated with $\text{Pref} \times D_t$ while the color indicates the statistical significance of the coefficient (p-value from two-sided t-test with heteroskedasticity-robust standard errors). The lighter the color, the closer the p-value is from zero. A grey color indicates a p-value greater than 0.2.

Figure C.11: *Stringency* and preferences (EVS/WVS)



(a) EVS-WVS (all)

(b) EVS-WVS (pre-2020)

Note: The figure illustrates 364 coefficients from a panel regression of economic preferences on the Stringency Index at day D_t ($t \in [1, 365]$) along with several control variables. The left panel displays estimates for each of the four preference measures obtained from all countries included in the EVS/WVS. The right panel display estimates for four preference measures obtained from the countries surveyed before 2020 in the EVS/WVS. Each regression includes all preference variables (*e.g.* risk-taking, patience, altruism or trust) interacted with the day dummies D_t as well as the (log) number of cases at date t , the day dummies and country-level fixed effects. The graph for each preference summarizes the information obtained from the regressions about the corresponding preference. The dot for date t conveys two pieces of information: the position indicates both the sign and magnitude of the coefficient associated with $\text{Pref} \times D_t$ while the color indicates the statistical significance of the coefficient (p-value from two-sided t-test with heteroskedasticity-robust standard errors). The lighter the color, the closer the p-value is from zero. A grey color indicates a p-value greater than 0.2.

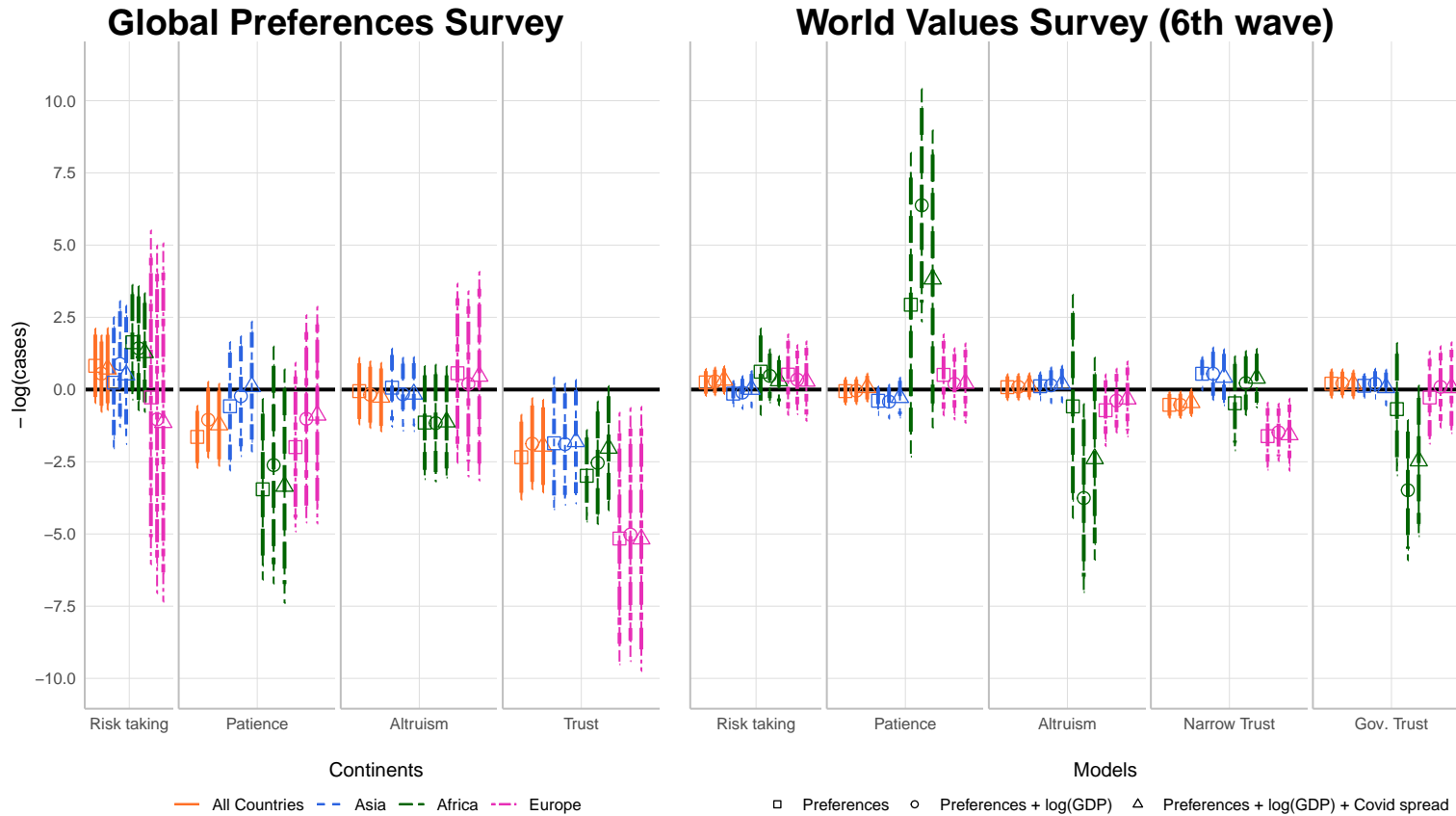
C.4.4 Analysis by continents

To investigate the possibility that our results are driven by some cluster of countries, we replicate our regression analysis separately for each continent. Considering that the number of countries included in our datasets per continent is limited and varies across continents, the obtained standard errors of the estimated coefficients are likely biased. As a consequence, our objective is to assess whether the signs of the coefficients associated with preferences are globally consistent, without a particular emphasis on their statistical significance. Furthermore, we do not include the full set of controls but restrict our attention to the nations' wealth and the international spread of the virus.

The figure C.12 shows that the sign of the relationship between preferences and *Responsiveness* indeed tends to be similar over continents, no matter whether controls are added or not. Our results are thus not conspicuously driven by a sample of countries (such as the wealthiest countries).

We observe more diverging results with respect to *Stringency*, especially for patience and altruism between Europe and Asia in the WVS dataset (figure C.14). Yet, considering that the results on the GPS are consistent and that the worldwide analysis provides highly-significant and consistent results makes us believe that we likely observe some minor regional discrepancy in the coefficients.

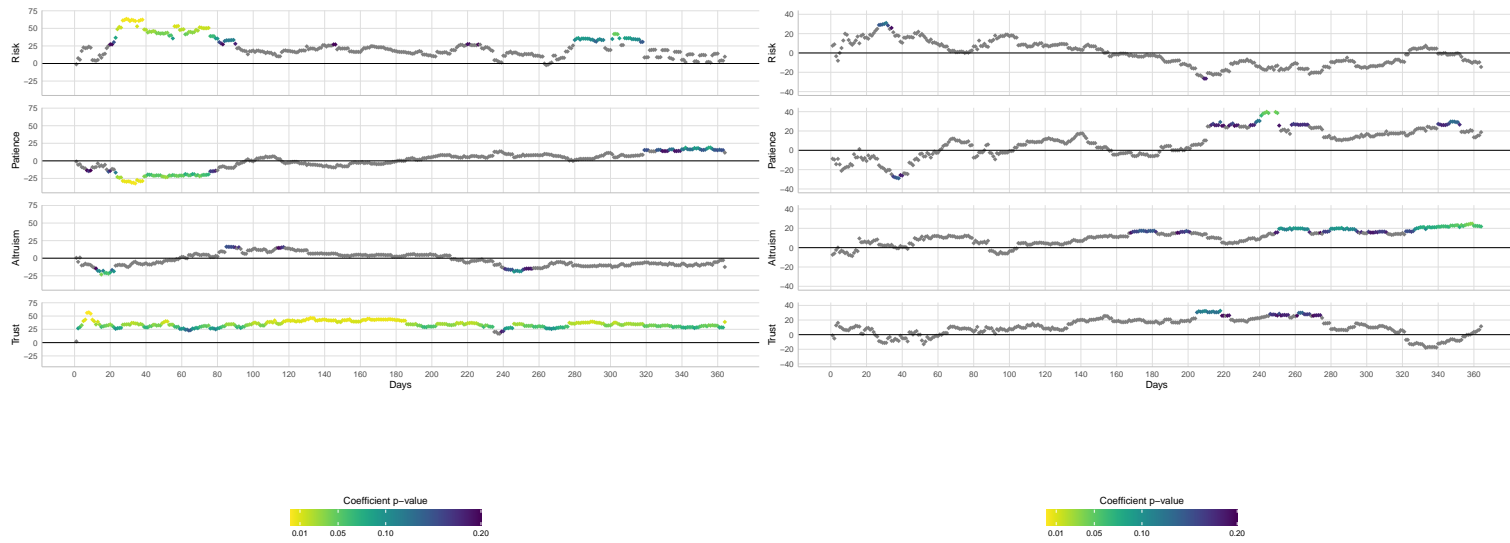
Figure C.12: *Responsiveness* over continents



Note: Each dot corresponds to the estimated coefficient associated with a given preference measure in OLS regressions of governments' *Responsiveness* on nationally-aggregated preferences. The bars correspond to 90% (shorter) and 95% (larger) confidence intervals, computed with robust standard errors. If one bar overlaps 0, the corresponding coefficient is not statistically significant at the given confidence level. The regressions include either only the preference measures ("raw preferences") or add some controls. The $\log(\text{GDP})$ variable is the (log) Gross Domestic Product per capita in the country and the *COVID spread* corresponds to the number of days between the first worldwide covid case and the first case in the country. The left panel displays countries surveyed in the GPS (75 countries), the right panel those taking part of the 6th wave of the WVS (59 countries). Due to severe issues with sample size, we do not report the full control regression on Northern-American countries.

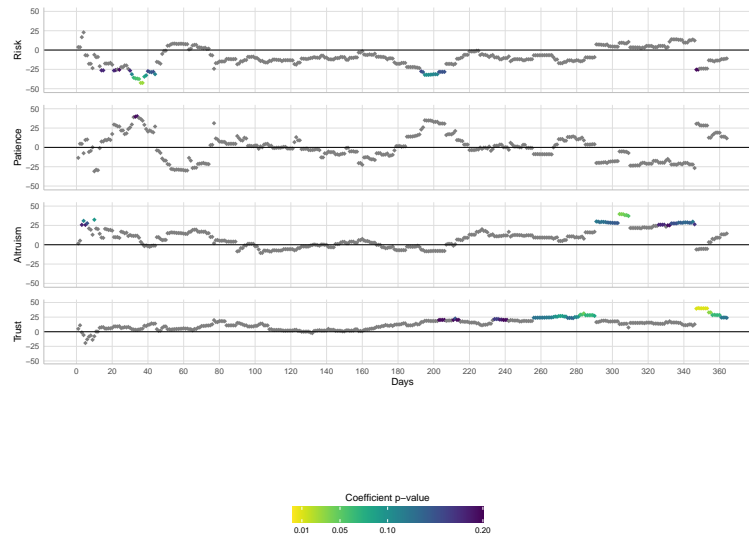
The figures C.13 and C.14 illustrate 364 coefficients from continent-level panel regressions of economic preferences on the Stringency Index at day D_t ($t \in [1, 365]$) along with several control variables using either the GPS dataset or the WVS dataset. Each regression includes all preference variables (*e.g.* risk-taking, patience, altruism or trust) interacted with the day dummies D_t as well as the (log) number of cases at date t , the day dummies and country-level fixed effects. The graph for each preference summarizes the information obtained from the regressions about the corresponding preference. The dot for date t conveys two pieces of information: the position indicates both the sign and magnitude of the coefficient associated with $\text{Pref} \times D_t$ while the color indicates the statistical significance of the coefficient (p-value from two-sided t-test with heteroskedasticity-robust standard errors). The lighter the color, the closer the p-value is from zero. A grey color indicates a p-value greater than 0.2.

Figure C.13: *Stringency* and preferences by continents (GPS)



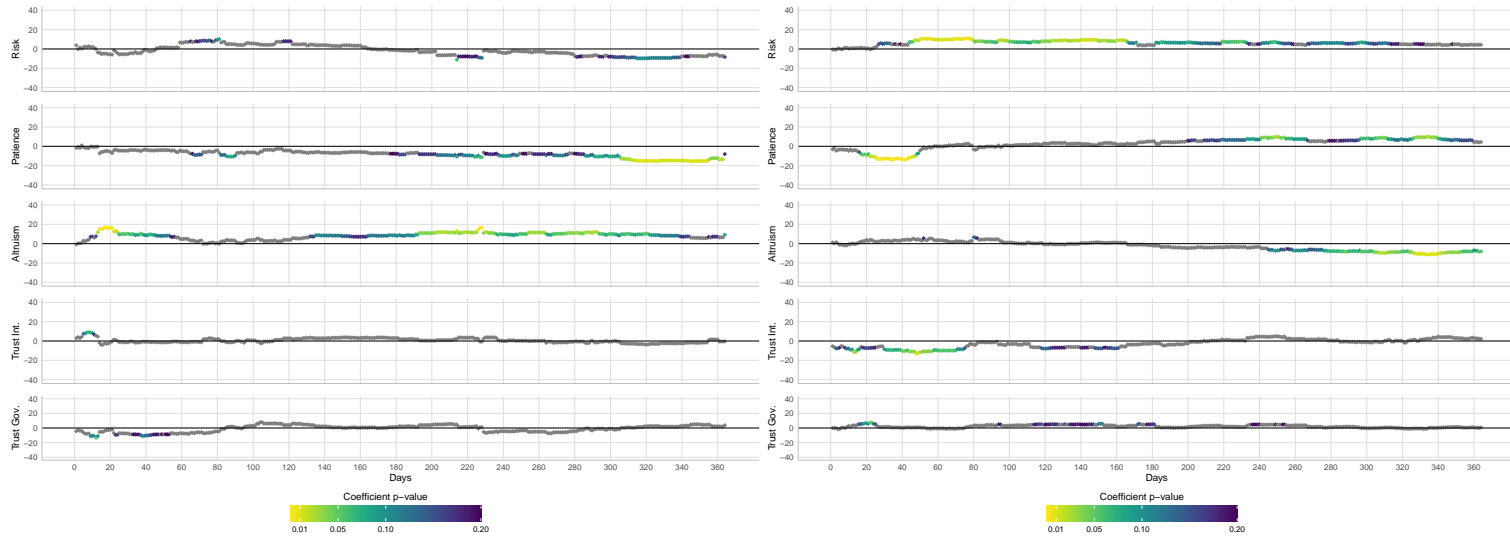
(a) Europe

(b) Asia



(c) Africa

Figure C.14: *Stringency* and preferences by continents (WVS)

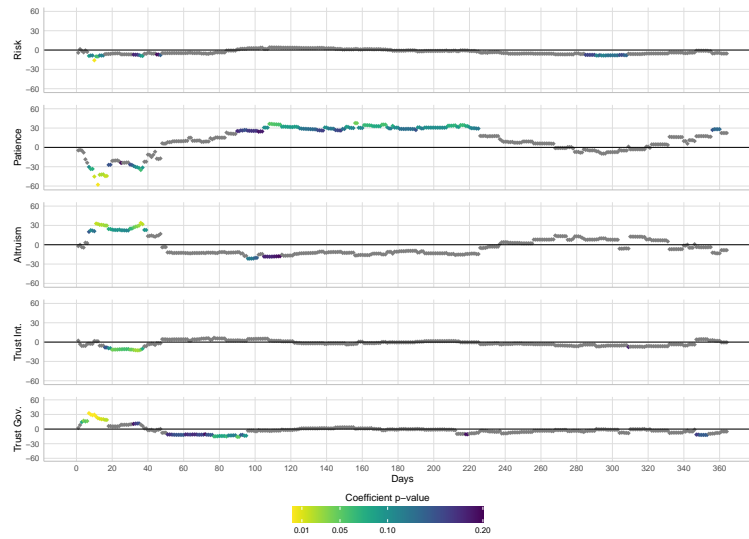


358

APPENDIX C

(a) Europe

(b) Asia



(c) Africa

References

- Alimohamadi, Y., Taghdir, M., and Sepandi, M. (2020). Estimate of the basic reproduction number for covid-19: a systematic review and meta-analysis. Journal of Preventive Medicine and Public Health, 53(3):151.
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. Games and Economic Behavior, 10(1):122–142.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. The Quarterly Journal of Economics, 133(4):1645–1692.
- Falk, A., Becker, A., Dohmen, T. J., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Working paper, Netspar discussion paper.
- Falk, A. and Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. Science, 362(6412).
- Gottlieb, C., Grobovšek, J., Poschke, M., and Saltiel, F. (2021). Lockdown accounting. The BE Journal of Macroeconomics.
- Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., et al. (2021). A global panel database of pandemic policies (oxford covid-19 government response tracker). Nature Human Behaviour, 5(4):529–538.

General Conclusion

Political decisions are increasingly nurtured by behavioral insights at all administrative levels.⁸ The general idea is to take into account the behavioral determinants of citizens' decisions (such as cognitive limitations or non-standard preferences) to design policies which will enhance citizens' welfare. At the same time, the scientific understanding of such political decisions predominantly assumes that political decision-makers are rational and self-interested. These two assumptions have fundamental implications both at the theoretical level and at the empirical level. For instance, they determine how politicians react to behavioral insights on public policy, such as regarding compliance to sanitary recommendations or vaccination uptake. This dissertation proposes to analyze political decisions from a behavioral perspective along three lines of inquiry.

The chapter 1 investigates whether politicians indeed are "rational", in the sense that they maximize some utility function. Specifically, we invited actual politicians among the French local pool to participate in an economic experiment during which participants were asked to make endowment allocations. Building on the equivalence between the axioms of revealed preference and the utility maximization framework, we then non-parametrically test the proximity of observed decisions with rational behavior using various rationality indices. Finally, we compare such indices between politicians and two control groups, a student control group and a non-student control group with similar socio-demographic characteristics as politicians, as well as among politicians, between a lab experiment and a field one. We find that politicians deviate from strict rationality but neither often nor by substantial amounts. Furthermore, we observe rationality indices to be close between politicians and a comparable group, although the former appears significantly less rational than a student group.

⁸For a recent example of local governments' interest in behavioral economics, see Stoll (2022).

GENERAL CONCLUSION

Finally, we do not find differences between politicians coming from the lab sessions or coming from the field sessions.

These results on the rationality of politicians' decision-making may be interpreted in two ways. On the one hand, they entail that the rational choice framework, which prevails for instance in public choice analyses, fit politicians' data rather successfully. As a consequence, dismissing such framework would probably be a mistake: it is useful to theorize on politicians' behavior and to make benchmark predictions as to how one may expect them to react in many different situations. On the other hand, as politicians do not make more efficient decisions than comparable citizens, our results imply that the inconsistencies which have been observed with non-politician samples are likely to prevail for politicians as well. Furthermore, as politicians are less rational than students in our experiment, cognitive limitations which have been observed with student samples may be exacerbated with politicians. Such observations would call for a better understanding of choice anomalies made by politicians and for a general reflexion on how to maximize efficiency in public decision-making which would incorporate behavioral insights at the level of the individual politician (e.g: how to nudge the nudger?).

The chapter 2 investigates whether politicians indeed are "self-interested", as opposed to being moved by alternative motivations, and especially distributional concerns. Using the same experimental data as in the chapter 1, we estimate a structural utility function which enables to finely disentangle two tradeoffs at the individual level: personal gains' maximization (Selfishness) versus maximization of other people's gains (Altruism) on the one hand, and minimizing inequalities (Equality) versus maximizing the sum of the gains (Efficiency) on the other hand. Such an empirical strategy allows to evaluate how politicians value the four behavioral motives in a simple decision environment which lays the emphasis on empirically-relevant tradeoffs in policy-making. We find that politicians' decisions are much more other-regarding and much more concerned about inequalities than the decisions made by both our control groups. Furthermore, we find little differences between lab and field sessions. Such findings coincide with the literature in other disciplines which emphasizes the importance of intrinsic motivations for politicians, yet in our study such motivations are directly inferred from incentivized choices in a tightly-controlled environment. These results have implications for the theoretical modelization of politicians' behavior, for the building of efficient incentives and in-

stitutions surrounding their daily activity, and for how scientifically-grounded policy recommendations could be more efficiently conveyed to the ears of the Prince.

Of course, such a study is ultimately only one element of a broader research program. In particular, it focuses on the individual decisions made by actual politicians in a very specific setting which omits important dimensions of the political decision-making process.⁹ In particular, it sets aside two important collective dimensions. The first collective dimension is that politicians decide for other citizens than themselves. This may directly affect the type of motivations that policy-makers would prioritize. Furthermore, these other persons in fact elect politicians in representative democracies, which means that citizens' preferences directly enter the decision-making process and compete with politicians' preferences.¹⁰ The second collective dimension is that politicians do not decide alone: they bargain and cast votes with their peers when designing a policy, and they decide based on interactions with many other actors (lobbies, administrative staffs, experts, etc.). Again, the interplay between such feature and politicians' individual preferences over budgetary choices could be an interesting research avenue, which ultimately connects to the large literature on (experimental) social choice.

In addition, our results underline some sort of primary taste for equalization among politicians, or an uniform distribution as a focal point in preferences, which may have important empirical implications. For instance, it may supply an explanation as to why some intermunicipal structures agree upon the principle of an equalization scheme among the partner municipalities, even though some municipalities obviously have a diverging interest and even though the scheme may turn out to be ineffective. In our view, scrutinizing the importance of such a focal point for politicians in other settings would be of particular interest. One may think of at least two follow-ups: replicating such findings in other decision environments (e.g., at higher administrative layers or in other countries) and investigating the link

⁹We acknowledge that this point also holds for the chapter 1.

¹⁰Investigating the importance of politicians' preferences for political decisions when citizens' preferences compete with them indeed is one major inquiry in political economy. Though disentangling whose preferences matter for political decisions with observational data may be difficult, it would certainly be possible to design an experimental study which would help address such representativity issue. In particular, one may consider a taxation/redistribution setting where the decision-maker's distributional preferences would likely play a role and where actual citizens could also express their preferences while varying the political institutions (e.g., representative democracy or dictatorship).

GENERAL CONCLUSION

between this taste for equalization and actual policy decisions, both theoretically and empirically. Finally, supplementing our empirical analysis with a theoretical framework which would help explain why there exists such a focal point, possibly inspired by the literature on political selection, would also constitute an important follow-up.

The results highlighted in both chapters 1 and 2 of this dissertation and their respective appendices also are useful from a methodological perspective. Our findings regarding the comparability of laboratory and field sessions with respect to both the decision-making process and the underlying preferences indeed appear a valuable finding for future experimental studies with actual politicians. Specifically, they entail that experimental data collected in the laboratory is likely to provide useful information even for politicians which may ultimately not participate in the experiment. Furthermore, our results on both our control groups and on socio-demographic variables complement existing findings in the literature. In particular, we apply recently-developped tests for homothetic preferences to a new dataset and recover previous results which were obtained with experimental data from past studies. We also document a link between political preferences and distributional concerns, even with subjects which are likely more politically-engaged than the remainder of the population and using different definitions of political preferences. Finally, we find interesting relationships between past experience in economic experiments and observed behavior with our student subjects, which would deserve some specific attention as it bears important implications for the interpretation of experimental results in general.

The chapter 3 adopts a different methodological stance than the one relied on in both chapters 1 and 2. Instead of collecting new data based on experiments to address specific issues related to political decision-making, it uses experimental data as observational data to underline stylized facts regarding actual political decisions. As a consequence, it is one way to directly look into the behavioral determinants of policy choices. Specifically, we make use of the fact that the COVID-19 epidemic was largely unexpected and quickly affected the entire world population to investigate the link between populations' risk, time and social preferences (altruism and trust) and the way governments reacted to the crisis. Building on large-scale datasets regarding both emergency sanitary policies around the world, which sub-

stantially modified populations' daily life, and populations' preferences before the virus stroke, we correlate governments' responsiveness and the stringency of the national response over time with such preferences. Our main finding is that behavioral motives do matter for policy-making, in the sense that societies with different preferences reacted differently to the epidemic, even when controlling for some important epidemic-relevant national characteristics and for the non-random spread of the virus. In particular, the trust level within the population appears to be crucial.

Obviously, such a correlational study cannot provide any conclusive elements regarding the causal link between political decisions and populations' preferences. Indeed, finding a counterfactual situation appears difficult considering the nature of the shock (a worldwide, sudden and lightning shock) and the level of analysis (country-level). Furthermore, we cannot disentangle the underlying mechanisms which may play a role in the relationships which we observe. For instance, we find that a higher trust level delays the first implemented restriction. However, we cannot state whether it is due to a lower demand for policy intervention or to the spontaneous adoption of disease-mitigation behavior on the part of the population which diminishes the need for such policy intervention. In addition, we cannot separate citizens' preferences from politicians' preferences, both of which are likely to play a role. Such results therefore ultimately calls for the incorporation of populations' behavioral preferences in our understanding of political decisions, and in particular for the building of theoretical models which would embrace behavioral insights.

Altogether, this dissertation seeks to contribute to a growing body of research which seeks to incorporate methods and analyses from behavioral and experimental economics to our understanding on the way political decisions are made empirically. In particular, we believe that a behavioral approach to political economy, especially on the supply side of policies, would provide important findings and help improve the functioning of democracies.

References

Stoll, S. (2022). L'économie comportementale aide à réfléchir aux outils d'intervention publique. Interview avec Nicolas Jacquemet (05/04). La Gazette des Communes. <https://www.lagazettedescommunes.com/797757/%e2%80%89leconomie-comportementale-aide-a-reflechir-aux-outils-dintervention-publique%e2%80%89>.

Stata References

This section contains references to the user-written Stata commands which we used in our analyses (while not being part of the 2014 Stata version):

- The BOOTTEST command by Roodman et al. (2019)
- The CMP command by Roodman (2011)
- The FPRANK command by BenMamoun (2006)

BenMamoun, M. (2006). Fprank: Stata module to compute two-sample fligner-policello robust rank order test. Statistical software components s456739, Boston College Department of Economics.

Roodman, D. (2011). Estimating fully observed recursive mixed-process models with cmp. *Stata Journal*, 11(2):159–206.

Roodman, D., MacKinnon, J., Nielsen, M., and Webb, M. (2019). Fast and wild: bootstrap inference in stata using boottest. *Stata Journal*, 19(1):4–60.

Titre : Evaluations empiriques des déterminants comportementaux des décisions publiques

Mots clés : Décision publiques, Elus, Préférences révélées, Comportement, Mesures expérimentales

Résumé : Les décisions quotidiennes des élus ont un impact substantiel sur le bien-être des populations. La compréhension des motivations et des mécanismes présidant à ces décisions est donc cruciale. L'analyse économique traditionnelle des décisions publiques s'appuie sur un cadre micro-économique fondé sur une hypothèse de rationalité (i.e. les individus prennent des décisions cohérentes) et une hypothèse d'opportunisme individuel (i.e. les individus agissent pour améliorer leur bénéfice personnel). Ces hypothèses ont néanmoins été largement questionnées par les travaux d'économie comportementale qui ont connu un gain d'intérêt croissant dans l'élaboration des politiques publiques. Cette approche n'a cependant été que peu mobilisée dans la compréhension des décisions des élus.

Sur la base d'expérimentations économiques de laboratoire et de terrain, cette thèse questionne d'abord les hypothèses de rationalité (maximisation d'utilité) et d'opportunisme individuel (égoïsme) avec une population d'élus locaux (chapitres 1 et 2). Les résultats indiquent que les élus sont aussi rationnels mais accordent davantage d'importance à autrui et à la réduction des inégalités qu'une population comparable. Le chapitre 3 évalue le lien entre des mesures comportementales des préférences (goût pour le risque, patience, altruisme et confiance) et la réactivité du gouvernement et l'intensité des politiques publiques dans le monde pendant la crise du coronavirus. Nous observons que les facteurs comportementaux sont fortement liés aux décisions publiques.

Title: Empirical Investigations on the Behavioral Determinants of Political Decisions

Keywords: Political decisions, Politicians, Revealed preferences, Behavior, Experimental measures

Summary: The daily decisions of politicians have a substantial impact on populations' well-being. Understanding the motivations and mechanisms behind these decisions is thus crucial. The standard economic analysis of public decisions builds on a microeconomic framework founded on a rationality hypothesis (i.e., individuals make decisions that are coherent) and a selfishness hypothesis (i.e., individuals act above all to improve their personal benefit). These assumptions have, however, been largely challenged by behavioral economics research, which has gained increasing interest in public policy-making. However, this approach has been little used to understand the decisions of politicians.

Based on laboratory and field economic experiments, this dissertation first evaluates the validity of the rationality (utility maximization) and individual opportunism (selfishness) hypotheses with local politicians (chapters 1 and 2). The results indicate that politicians are as rational as a comparable population, but also value others and reducing inequalities more. Chapter 3 determines the relationship between behavioral measures of preferences (taste for risk, patience, altruism and trust) and both government responsiveness and the stringency of implemented policies around the world during the coronavirus crisis. We observe that behavioral factors are strongly correlated with political decisions.