



**HAL**  
open science

# Who speaks out when politicians tune in? Interpersonal trust and participation in public consultations

Clément Herman

► **To cite this version:**

Clément Herman. Who speaks out when politicians tune in? Interpersonal trust and participation in public consultations. Economics and Finance. 2020. dumas-03045175

**HAL Id: dumas-03045175**

**<https://dumas.ccsd.cnrs.fr/dumas-03045175>**

Submitted on 7 Dec 2020

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

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



**MASTER THESIS N° 2020 – 02**

**Who speaks out when politicians tune in ?  
Interpersonal trust and participation in public consultations**

**Clément Herman**

**JEL Codes: D70, D72, D90, Z10**

**Keywords:**



*Master's Thesis*

Who speaks out when politicians tune in?  
Interpersonal trust and participation in public consultations

September 3, 2020

**Clément Herman**

Ecole Normale Supérieure-PSL & Paris School of Economics

clement.herman@ens.fr

*under the supervision of*

**Ekaterina Zhuravskaya**

**Abstract**

Public consultations, especially online, are more and more commonplace in contemporary democracies. This work aims to study how interpersonal trust determines participation in public consultations, and the forms thereof. I use the *Grand Débat National*, that occurred in France in 2019, as a natural experiment to study the geographic cross-sectional relationship between a novel synthetic measure of local interpersonal trust and various aspects of participation in this public consultation. Then I study evidence from a small-scale experiment on interpersonal trust and participation in an online public consultation. I conclude that high-trust individuals participate more to public consultations compared to low-trust individuals, both at the extensive and intensive margins, and that there exists a discrepancy in their forms of participation, in terms of content, but also when it comes to textual or lexical aspects of participation. This result calls for questioning the common presumption that public consultations will give a better representation to and help design better policies for everyone, particularly low-trust individuals. In fact, it appears that public consultations can be instead captured by the demands of high-trust individuals.

*I would like to thank Ekaterina Zhuravskaya for invaluable advice and guidance as my supervisor, as well as Daniel Cohen for accepting to be the referee. I am deeply grateful to Stefanie Stantcheva for her support and inspiration, and also to Alessandra Casella for her inestimable backing for the experiment.*

JEL codes : D70, D72, D90, Z10

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Geographic cross-sectional relationship between trust and participation to the <i>Grand Débat National</i></b>	<b>6</b>
2.a	Context : the <i>Gilets Jaunes</i> movement and the set-up of the <i>Grand Débat National</i> . . . . .	6
2.b	Data and empirical strategy . . . . .	8
2.b.1	<i>Grand Débat National</i> data . . . . .	8
2.b.2	Other data . . . . .	9
2.b.3	Unit of analysis and data construction . . . . .	9
2.b.4	Econometric specification . . . . .	10
2.c	The synthetic local interpersonal trust variable . . . . .	11
2.d	Local interpersonal trust and participation to the <i>Grand Débat National</i> . . . . .	13
2.d.1	Participation variables of interest . . . . .	13
2.d.2	Cross-sectional results . . . . .	15
2.e	Local interpersonal trust and lexical aspects of online contributions . . . . .	16
2.e.1	The dictionary-based approach with the LIWC software . . . . .	16
2.e.2	Lexical variables of interest . . . . .	16
2.e.3	Cross-sectional results . . . . .	17
2.f	Local interpersonal trust and profiles of online contributors . . . . .	19
2.f.1	Construction of the profiles of respondents to the short questionnaires . . . . .	19
2.f.2	Construction of the profiles of respondents to the “Democracy and Citizenship” long questionnaire . . . . .	22
2.f.3	Cross-sectional results . . . . .	24
2.g	Robustness checks . . . . .	25
2.h	Keyness analysis : local interpersonal trust and lexical content of online contributions . . .	26
<b>3</b>	<b>Experimental evidence on interpersonal trust and participation to public consultations</b>	<b>28</b>
3.a	Purpose of the experiment and related literature . . . . .	28
3.b	An experiment on interpersonal trust and participation to an online public consultation . .	28
3.b.1	Experimental design . . . . .	28
3.b.2	Tested hypotheses . . . . .	30
3.b.3	Data collection . . . . .	30
3.c	Descriptive statistics . . . . .	31
3.c.1	Trust game variables . . . . .	31
3.c.2	Variables of participation to the questionnaire . . . . .	33
3.d	Results . . . . .	35
3.e	Remarks and limitations . . . . .	36
<b>4</b>	<b>Conclusion</b>	<b>37</b>
<b>5</b>	<b>Appendices</b>	<b>39</b>
5.a	List of questions from Grand Débat National . . . . .	39
5.a.1	Short questionnaires . . . . .	39
5.a.2	Long questionnaires . . . . .	42
5.b	Spatial micro-simulation : method, construction, and validity tests . . . . .	47

5.b.1	Method : synthetic micro-simulation using the iterative proportional fitting . . . . .	47
5.b.2	Construction of the local interpersonal trust variable with synthetic micro-simulation	48
5.b.3	Internal validation . . . . .	50
5.b.4	External validation . . . . .	51
5.b.5	Decomposition of the synthetic local interpersonal trust variable . . . . .	53
5.b.6	Modalities of the linking variables . . . . .	55
5.b.7	Summary statistics tables . . . . .	56
5.c	Building profiles of respondents using Latent Dirichlet Allocation . . . . .	57
5.c.1	The LDA algorithm to build profiles of respondents to the short questionnaires . .	57
5.c.2	Construction of the profiles of respondents to the “Democracy and citizenship” long questionnaire . . . . .	58
5.c.3	Profiles built from LDA analysis of the short questionnaires : top 15 answers per profile . . . . .	59
5.c.4	Profiles built from LDA analysis of the “Democracy and Citizenship” long ques- tionnaire : top 15 answers per profile . . . . .	62
5.d	Conditional scatter plots for the geographic cross-sectional analysis . . . . .	63
5.d.1	Participation variables . . . . .	63
5.d.2	Lexical variables . . . . .	65
5.d.3	Profiles of respondents variables . . . . .	67
5.e	Robustness check - Regression results with disaggregated local interpersonal trust variable	68
5.f	Keyness analysis . . . . .	75
5.f.1	High-trust and low-trust localities . . . . .	75
5.f.2	Text preprocessing . . . . .	75
5.f.3	Construction of the keyness statistic . . . . .	76
5.g	Appendix for the experiment on trust and participation to online participatory democracy	77
5.g.1	Figures . . . . .	77
5.g.2	First part of the experiment : trust game instructions . . . . .	78
5.g.3	Second part of the experiment : questionnaire on the economic consequence of the coronavirus pandemic . . . . .	80
5.g.4	Robustness check . . . . .	81

# 1 Introduction

In the last decades, participatory democracy practices have become more and more widespread in many countries around in the world (Pateman, 1970, 2012), and in particular in France (Blondiaux, 2009; Blondiaux, 2017). At first try, participatory democracy can be defined as the involvement of citizens into the political process outside of elections : citizens interact with elected representatives to define, design, and review policies, instead of letting elected representatives decide by themselves on behalf of citizens. It thereby combines elements of direct democracy and representative democracy. Pateman (1970) thus defines participatory democracy as a political model in which the maximum input (or participation) is demanded from the citizen.

Public consultations are one of the most popular forms taken by participatory democracy in recent years (Fishkin, 2011). Public consultations are the active process through which politicians, at the local or national levels, seek to obtain the public's input on policy-related issues. They offer a space in which citizens can express themselves to influence politicians and policies. They can take various forms : public meetings, voluntary surveys, whether online or not, suggestion boxes, citizen working groups... Public consultations are typically organized by local authorities for urban planning matters, especially when it comes to industrial sites, but also for the provision of many public goods such as schools, sport facilities etc. In France, there is a distinct tradition of this form of participatory democracy, known as *Débat public* (Public Debate) (Revel et al., 2007), embodied by the *Commission Nationale du Débat Public* (National Commission on Public Debate), set up in 1995. One notable past experience of a large scale implementation of a *Débat Public*, reviewed in Fourniau et al. (2019), is the *Débat National sur l'Avenir de l'Ecole* (National Debate on the Future of School) set up in 2003. With the rise of the Internet, more and more public consultations take place online, and can easily be organized at the national level. In France for instance, online public consultations have been organized in the past few years on the pension system reform, and the *revenu universel d'activité* (universal basic income), among others. One of the largest public consultation ever organized was actually the online public consultation by the European Union on seasonal time change which attracted 4.6 million respondents in 2018.

Aiming to tackle the estrangement of citizens from the political process, as well as provide them with an extra channel to influence policies, public consultations seek to attract participation from as many citizens as possible, and capture their analyses, experiences, and sentiment in a novel way. In particular, a commonly-held view, guiding the implementation of public consultations, posits that they give a voice to people who tend to not participate to traditional forms of political participation, such as the vote. However, public consultations always fall short of the representativeness which is usually achieved in opinion polls, because typically, only these individuals who care about a subject will participate in these consultations. Ultimately, political and social inequalities are often reflected in the participation in these consultations (Blondiaux, 2017). Nevertheless, with the advent of social media and the ever-expanding access to the Internet, public consultations today, especially online, are supposedly more easily accessible and publicized. Still, little is known about the kinds of public that do participate in these new forms of consultation. In particular, little quantitative work has been conducted, on recent instances of public consultations, about the workings and mechanisms of citizens' involvement into public consultations, especially online, and the self-selection of citizens that these consultations promote.

Since the pioneering work of Putnam (2000), interpersonal trust has attracted attention as a key factor driving behaviours and social phenomena. It has been shown to be a crucial determinant for economic processes (Zak and Knack, 2001; Aghion et al., 2010), and but also to explain the rise of populist politics (Algan et al., 2017; Dal Bó et al., 2018). In the contemporary French context, Algan et al. (2018, 2019a) study how low interpersonal trust is a decisive factor explaining the evolutions of the political preferences of the electorate, and in particular the divide between the radical left and the populist right. More recently, Algan

et al. (2019b) explore the *Gilets Jaunes* social movement of 2018-2019, and highlights how extremely low levels of interpersonal trust is a notable characteristic of this social movement's members and supporters.

In this work, I will examine how interpersonal trust can be mobilized as a key factor explaining participation in public consultations, and show that higher levels of interpersonal trust are associated with increased participation in public consultations. This increased participation takes place at the extensive margin, whereby more trusting individuals take more part in public consultations than less trusting individuals. But also, I will show that increased participation also takes place at the intensive margin : more trusting individuals take advantage of public consultations more fully and contribute at greater length than less trusting individuals. This double mechanism leads to public consultations being, metaphorically speaking, "taken over" by high-trust individuals.

On top of looking at the mere participation, I will also explore how interpersonal trust impacts what is expressed in these public consultations, by focusing on the forms of participation as well as the content thereof. It appears that interpersonal trust drives a discrepancy in the forms of expression to public consultations. Said otherwise, low-trust individuals do not tend to express themselves similarly to high-trust individuals in public consultations. Because of the higher presence of more trusting individuals, this leads to the overrepresentation of their language and demands in public consultations, relative to others, hence a potential capture of public consultations by these more trusting individuals.

The ideal setting to study this question would have been one where individual micro-data on actual participants to public consultations would have been available, and in which their level of interpersonal trust would be evaluated. In the absence of such data, I provide two sets of evidence to highlight the role of interpersonal trust as a key factor explaining participation in public consultations and the forms thereof.

- First, I mobilize the *Grand Débat National*, a large-scale public consultation that happened in France in the early months of 2019, as a natural experiment, and set up a geographic cross-sectional approach that studies the relationship between a novel synthetic measure of local interpersonal trust and various measures of participation to the *Grand Débat National*, offline and online, at the intensive and extensive margins. I also study the relationship between interpersonal trust and the forms and content of participation to public consultations by employing natural language analysis and machine learning tools to generate some tractable metrics. Finally, I use the keyness approach to study more specific textual aspects of the contributions to this public consultation.
- Then, I present a small-scale experiment that has been conducted to provide additional evidence on the relationships uncovered in the geographic cross-sectional approach. The experiment makes use of an alternative way of measuring interpersonal trust, through the trust game put forward in Berg et al. (1995). The experiment further supports the claim that interpersonal trust drives participation to online public consultations, especially at the intensive margin.

The relevance of this work is twofold. First, because it helps shed light on the mechanisms of self-selection into public consultations, and on how interpersonal trust is an interesting lens to study the overrepresentation of the demands of certain segments of the population relative to others. These findings should lead to extra caution by policy-makers when launching and interpreting the results of public consultations, by giving a better understanding of the biases of such participatory democratic tools. In particular, in the case where there is indeed a gap in the demands expressed by low-trust and high-trust individuals, and if the government naively applies the conclusions from public consultations, it would likely overlook the input from low-trust individuals who tend to be relatively absent from public consultations. Consequently, public consultations would not fulfill their promises of giving all citizens an extra channel to influence policy-making, but may actually aggravate inequalities in influence over policies.

Its relevance is even greater in the case of France, due to the specific context in which the *Grand Débat National* took place. As highlighted in [Algan et al. \(2019b\)](#), participants to the *Gilets Jaunes* social movement were characterized by low levels of interpersonal trust. The *Grand Débat National*, organized in the wake of this movement, aimed at channeling the social movement into a novel yet institutionalized form of political action. Therefore, the finding that the *Grand Débat National* was actually taken over by high-trust individuals, indicates that the government did not manage to reach out to the *Gilets Jaunes* protestors with this public consultation.

Section 2 provides the cross-sectional geographical evidence, based on the *Grand Débat National* natural experiment, it also describes the data used, methods, and results. Section 3 presents the experiment on interpersonal trust and participation to online public consultation. Section 4 concludes.



## 2 Geographic cross-sectional relationship between trust and participation to the *Grand Débat National*

In this section, I mobilize the *Grand Débat National* as a natural experiment to study the relationship between interpersonal trust and participation to public consultations, and the forms thereof. Subsection 2.a presents the historical context of the *Grand Débat National*. Subsection 2.b presents the data used as well as the empirical strategy for the rest of this section, except subsection 2.h. Subsection 2.c presents the novel synthetic measure of interpersonal trust at the local level, which is the key explanatory variable of interest mobilized throughout the section. The outcome variables of interest used to measure diverse aspects of participation to *Grand Débat National* are presented in subsections 2.d (participation variables), 2.e (lexical variables), and 2.f (profiles of respondents). Each of these three subsections also presents the results of the econometric specification. Additional robustness checks of this empirical strategy are presented in subsection 2.g. Finally, subsection 2.h presents a different empirical strategy, the keyness analysis, which provides some insightful results on the lexical diversity of answers across life zones with different levels of interpersonal trust.

### 2.a Context : the *Gilets Jaunes* movement and the set-up of the *Grand Débat National*

In October 2018, following the decision by the French government to increase the taxes on fuel, justified by environmental concerns, a social movement emerged, aiming to repeal this increase. Very soon, this social movement gained momentum, both online (with online petitions and social network activity, studied by [Boyer et al. \(2020\)](#)) and offline (with road blockades, especially at roundabouts). The demands went from the repeal of the fuel tax to demands for more social and economic justice, as well as institutional reforms to introduce direct democratic tools (citizen's initiative referendums). Horizontal and leaderless, the movement and its members soon developed their own distinctive eponymous sign : the Yellow Vest ("*Gilet Jaune*"). In November and December 2018, the movement reached an apex, with many demonstrations all over France, and some violent clashes with police forces in Paris especially.

After weeks of protests and blockades across the country, President Macron announced, in December 2018, that a *Grand Débat National* (Great National Debate, from now on referred to as GDN) would be launched in 2019. It was meant to be large-scale public consultation, aiming to engage citizens and collectively find a way out of the social movement. President Macron announced that it would focus on four topics : ecological transition, fiscal issues, democracy and citizenship, the organization of the State and of public services. These topics were indeed identified as being the key domains that were driving the discontent and fueled the *Gilets Jaunes* movement.

The GDN took place in various forms :

- 16337 *cahiers de doléances* (grievance books), set up by mayors in townhalls.
- 10134 local meetings, organized by citizens, elected representatives, charities or companies. Local meetings could be announced on the official GDN website, and participants could also hand in a report on the website.
- An online platform was set up for citizens to answer directly to a set of questions designed by the government on these four topics, it was accessible from the GDN official website.
- 4 national thematic conventions gathered unions, employer organizations and associations.
- 21 regional citizen conventions, which gathered randomly selected citizens from a given administrative region, under the model of citizens' assemblies.

- Some citizens also contributed via mail and email (27374 contributions).

With official numbers indicating more than 2.8 million people who consulted the online webpage of GDN, 1.2 million participants, shared between 500 000 online, 500 000 in local citizen meetings, and 200 000 in *Cahiers de doléances*, the government claimed the GDN was a success. [Fourniau et al. \(2019\)](#) independently reviewed and confirmed these figures.

The government requested various private companies (consultants, data science, and polling companies) to provide syntheses for the diverse and tremendous amount of material collected. The GDN came to a close with a press conference of President Macron on April 25. Several new policies were announced, based on the syntheses.

The GDN was characterized by a strong emphasis on the transparency in the data collected. All data on online contributions and local meetings reports was easily accessible on the official website during the GDN. The *Cahiers de Doléances* are still under the process of numerisation by the French National Library. Similarly, all syntheses are available online.

So far, GDN has been studied from several perspectives. No joint analysis of offline and online activity has been carried out, however, concerning offline GDN activity, [Fourniau et al. \(2019\)](#) synthesizes the findings of the *Observatoire des Débats*, a grouping of researchers that studied GDN local meetings through a large-scale survey questionnaire administered to participants to randomly selected local meetings. Their main finding is that the sociology of participants to these local meetings is almost opposite to that of participants to the *Gilets Jaunes* movement : participants to the local meetings tended to be male, pensioners, educated, and relatively well-off. They also conclude on the great degree of similarity in political attitudes between participants and the Macron electorate in the 2017 elections.

[Ploux et al. \(2020\)](#) analyze the lexical content of GDN online contributions. They use advanced natural language analysis tools to study the underlying demands and topics mentioned by respondents. [Bellet et al. \(2020\)](#) also undertake semantic and lexical analysis of online contributions to the GDN platform. Their goal is to try and reproduce the syntheses made by the private companies requested by the government, and they conclude on the impossibility to reproduce their results, and the variability of conclusions depending on the tools used.

[Rouban \(2019\)](#) presents an in-depth study of a small subset of online contributions to GDN on the topic of democracy and citizenship, which underlines the tradeoff between trying to analyze comprehensively all the contributions and trying to analyze their content in-depth.

More closely related to our approach and mobilizing quantitative tools, [Bennani et al. \(2019\)](#) examine the local determinants of online participation to GDN. They take advantage of the geographic structure of the data, and highlight the role of the median standard of living and the level of education as key drivers of participation. They also draw some contrast between the different topics tackled by the different questionnaires. However, they fail to take into account the richness of the spatial diversity by relying only on the comparison at the *département* level. I will borrow from their cross-sectional geographic approach but will study a wider range of measures of GDN participation and consider a finer geographical unit of analysis.

## 2.b Data and empirical strategy

### 2.b.1 *Grand Débat National* data

All GDN data used here is in open access from the GDN’s official website<sup>1</sup>. The data used considers two forms of GDN activity : offline local meetings, and online contributions to the platform. Unfortunately, data on grievance books would have been fit for geographic cross-sectional analysis but it was not yet available.

To study offline GDN activity, i.e. the local meetings, two datasets are used : the dataset on the local meetings which were announced on the dedicated website, and the dataset on the local meetings for which a report was uploaded on this website. For both, information is available on the date and localization of the event. The reports also include an estimate of the number of participants. Some meetings were announced but no report was filed, and conversely some meetings were reported but not announced on the website beforehand. I define as “local meeting” any such event that happens at a given date and in a given zipcode, whether it was announced on the website and/or whether a report was filed afterwards. Based on this definition, 9061 local meetings are considered for the analysis. To estimate the average number of participants, I rely on the subset of 5295 meetings for whom an estimate on the number of participants was given.

Data on online activity consists in the answers given to eight online questionnaires. For each of the four topics of the GDN, two questionnaires were set up, a short one and a long one. Each respondent to the online platform was first asked to choose a topic, and then the type of questionnaire (long or short). Each respondent was free to participate to as many questionnaires as wanted. Within each questionnaire, respondents were able to skip questions if they wanted to. Table 1 shows that long questionnaires attracted on average less respondents than short ones. Questionnaires on the topics “Ecological transition” and “Fiscality and public spending” attracted more respondents than other topics. It is not a surprise that long questionnaires attracted much more answers than short questionnaires, since the short questionnaires consisted only of close-ended questions, and was meant to be answered quickly. On the other hand, the long questionnaires included both close-ended and open-ended questions and were thus much time-consuming and demanding. For the complete list of questions and classification into open and close-ended, see appendix section 5.a.

Type of questionnaire	Democracy and citizenship	Ecological transition	Fiscality and public spending	Organisation of the state and public services
Long	99 229	130 790	148 545	93 458
Short	326 467	342 154	334 698	325 252

Table 1 – Number of contributors to each online questionnaire, by topic and type of questionnaire

After having answered to a questionnaire, respondents were asked to give out an email address. These email addresses were anonymized, but enabled to identify respondents who answered to multiple questionnaires. The cleaning process involved the removal of duplicates, i.e. questionnaires who were taken multiple times by the same respondent.

Respondents were also asked about their identity : whether they were a citizen, an elected representative or institution, a non-profit organization, or a for-profit organization. For the sake of consistency and comparability of answers, I selected only contributions from self-described citizens, which make up 84% of contributors.

<sup>1</sup><https://granddebat.fr/pages/donnees-ouvertes>

Finally, respondents to the online questionnaire were asked to report the zipcode of their current place of residency. Respondents with invalid zipcodes were discarded. 409919 respondents are eventually in the final dataset.

No extra individual information was collected on the respondents. The lack of any additional individual micro-data collected on the online participants, deplored by [Fourniau et al. \(2019\)](#), calls for the choice of a geographic cross-sectional analysis where respondents are gathered at a certain geographical unit, and where it is these geographical units that are studied, and not individuals themselves.

### 2.b.2 Other data

I use municipality-level data from INSEE’s census data from 2016 (distribution of age, gender, and level of diploma) and total population as of 2019, as well as the 2017 presidential election results from French *Ministère de l’Intérieur*. Correspondance tables from INSEE are used to account for the change in municipalities’ boundaries over the period considered and harmonize it with the current boundaries as of 2020. 34478 municipalities are considered, by restricting the analysis to Metropolitan France, Corsica excluded.

### 2.b.3 Unit of analysis and data construction

I choose INSEE’s “life zone” (*bassin de vie*) as the unit of analysis. A life zone groups several municipalities, and is defined as the smallest territory on which the inhabitants have access to the most current equipment and services. 1663 life zones have been defined by INSEE in 2012, gathering 21 municipalities on average. The analysis is restricted on the 1632 life zones of metropolitan France, Corsica excluded.

The choice of this unit of analysis borrows from [Boyer et al. \(2020\)](#), as well as [Algan et al. \(2019c\)](#) who both use this unit of analysis to study the *Gilets Jaunes* movement. The choice of this level of aggregation is also made because a smaller level of aggregation, such as the zipcode (5944 units in metropolitan France), is unattractive due to the absence of a straightforward match between municipalities and zipcodes, and the existence of many municipalities spanning over several zipcodes. On top of that, this level of aggregation has boundaries that match purportedly the citizen’s daily lives, which is particularly relevant when considering participation to GDN local meetings.

The aggregation of data at the life zone level is straightforward in the case of municipality-level data, since a life zone is a grouping of municipalities. In the case of zipcode-level data such as GDN data, I use correspondance tables from INSEE between zipcodes and municipalities, to weigh zipcode-level data between one or several life zones. The weighting is performed in order to avoid double-counting issues. Life zones differ a lot in their size, populations, and population densities. Indeed, a life zone can cover a small rural territory, but a single life zone also covers most of the Paris region (see table 2 for summary statistics on life zones’ populations, sizes, and population densities).

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Surface (in hectares)	1,632	33,088.840	28,489.650	936	14,926	26,256.5	42,757	396,046
Total population as of 2019	1,632	40,134.030	287,912.700	1,911	9,740.8	14,958	24,618.2	11,019,994
Population density (inhabitants per hectare)	1,632	1.118	1.507	0.059	0.376	0.688	1.263	27.825
Log-transformation of population density	1,632	-0.339	0.909	-2.832	-0.978	-0.374	0.233	3.326

Table 2 – Summary statistics of demographic variables for life zones (INSEE’s *bassin de vie*)

## 2.b.4 Econometric specification

To investigate the role played by interpersonal trust in driving participation to GDN, as well as the forms taken by this participation, I study the geographic cross-sectional relationship between a novel synthetic measure of local interpersonal trust and a number of variables capturing diverse aspects of participation to GDN, at the life zone level.

This empirical strategy borrows directly from [Algan et al. \(2019b\)](#) in their analysis of the geographic determinants of the *Gilets Jaunes* movement : they look for a correlation between local synthetic well-being and trust variables and the declared roundabout blockades. The choice of this empirical strategy also stems from the limitations of survey data : in their case, individual survey data can only capture declared support for the *Gilets Jaunes* movement, but not actual participation to the movement, while here, individual micro-data on GDN participation is simply not available.

Extra variables are added to the regressions as robustness checks : *département*<sup>2</sup> fixed effects and the log-transformation of the population density capturing the rural-urban divide that both could drive the observed relationships. A log-transformation is used to deal with the outlier issues seen above. On top of that, standard errors are clustered at the *département* level to account for spatial autocorrelation.

From the very nature of the synthetic local interpersonal trust variable (see section 2.c), no extra socio-demographic or political control is added to the regression. In effect, this synthetic variable is akin to a linear combination of various socio-demographic and political local determinants. As a result, only correlational evidence can be presented in the section, i.e. the correlation between the synthetic local interpersonal trust variable and the variables of GDN activity, controlling for *département* fixed effects and population density.

The econometric specification is thus

$$Y_{i,d} = \alpha_{i,d} + \beta IPT_{i,d} + \delta_d + \gamma D_{i,d} + \epsilon_{i,d}$$

with  $Y_{i,d}$  the outcome variable of interest capturing an aspect of participation to GDN, in life zone  $i$  that belongs to *département*  $d$ .  $IPT_{i,d}$  is the local interpersonal trust in life zone  $i$ ,  $\delta_d$  is the *département* fixed effect, and  $D_{i,d}$  is the log-transformation of population density.

---

<sup>2</sup>*Départements* are French local districts, there are 94 of them in Metropolitan France, Corsica excluded.

## 2.c The synthetic local interpersonal trust variable

I generate a synthetic local interpersonal trust variable using spatial microsimulation. More specifically, I use the iterative proportional fitting procedure described in [Lovelace and Dumont \(2016\)](#). This method offers a synthetic micro-founded way to generate detailed geographical data on interpersonal trust that is otherwise unavailable.

Mapping out the geographic distribution of interpersonal trust is not a new question. [Alesina and La Ferrara \(2002\)](#) use the reported zipcodes of respondents to a large-scale survey to map out the distribution of interpersonal trust across different states in the US. To measure interpersonal trust, they use the traditional question “Generally speaking, would you say that most people can be trusted, or that you can’t be too careful in dealing with others?”. However, this solution is not reproducible in this context because of the absence of sufficiently detailed geocoded microdataset that would enable to approach interpersonal trust at the life zone level. This calls for an alternative approach to measure (or at least proxy for) interpersonal trust at a local level.

[Algan et al. \(2019b\)](#) propose an alternative method to map out interpersonal trust at the local level in France. The chosen approach consists in using survey data to estimate a linear relationship between interpersonal trust, measured with the same question as above, and some individual economic, social and political variables. The estimated coefficients are then extrapolated to predict local interpersonal trust using the equivalent variables at the local level. Spatial microsimulation offers an approach which is similar in spirit, yet more robust and micro-founded, to map out interpersonal trust at the local level.

Spatial microsimulation defines a set of tools that aim to generate some spatial microdata by combining survey microdata and geographic aggregated datasets. While it offers multiple methods for implementation, such as iterative proportional fitting and combinatorial optimization, I rely on iterative proportional fitting, which is a reliable and tractable method to generate spatially disaggregated data ([Wong, 1992](#)). Spatial microsimulation has had many applications : [Edwards and Clarke \(2009\)](#) uses it to estimate the rates of obesity at a local level in Leeds (UK), [Tomintz et al. \(2008\)](#) generate the distribution of smokers in the same city, while [Lovelace et al. \(2014\)](#) and [Moeckel et al. \(2003\)](#) use this method to study commuting patterns and travel demand. More closely related to our research question, spatial microsimulation has been used in [Hermes and Poulsen \(2013\)](#) to map out the distribution of interpersonal trust of Sydney (Australia), at a very detailed geographical level. On top of geographic applications to generate spatially disaggregated data, microsimulation is also used in the EUROMOD model ([Sutherland and Figari, 2013](#)) to simulate the impacts of taxes and benefits on household income in the European Union.

Appendix section 5.b presents in detail the spatial microsimulation method, and how it is performed here to create a local synthetic interpersonal trust variable at the life zone level. It also presents internal and external validity checks, as well as a decomposition of the local synthetic interpersonal trust variable into a linear combination of socio-demographic and political variables. The assumptions of the method, as well as the data used, are thoroughly discussed.

In short, synthetic microsimulation consists in creating a synthetic population for every life zone, by combining survey microdata and geographic datasets on populations at the life zone level. In the survey microdata, the level of interpersonal trust is known for every individual, and is measured with the question : “On a 0 to 10 scale, would you say that you can trust most people or that one is never cautious enough when treating with others?”. The synthetic local interpersonal trust variable corresponds to the average value for this variable in the synthetic population at the life zone level. While answers can individually span from 0 to 10, when averaged at the life zone level, there remains some variation in average values. The synthetic local interpersonal trust variable is mapped out in figure 1, while table 3 presents some summary statistics of this variable.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Local interpersonal trust	1,632	3.960	0.124	3.639	3.871	3.953	4.039	4.472

Table 3 – Summary statistics of the local interpersonal trust variable

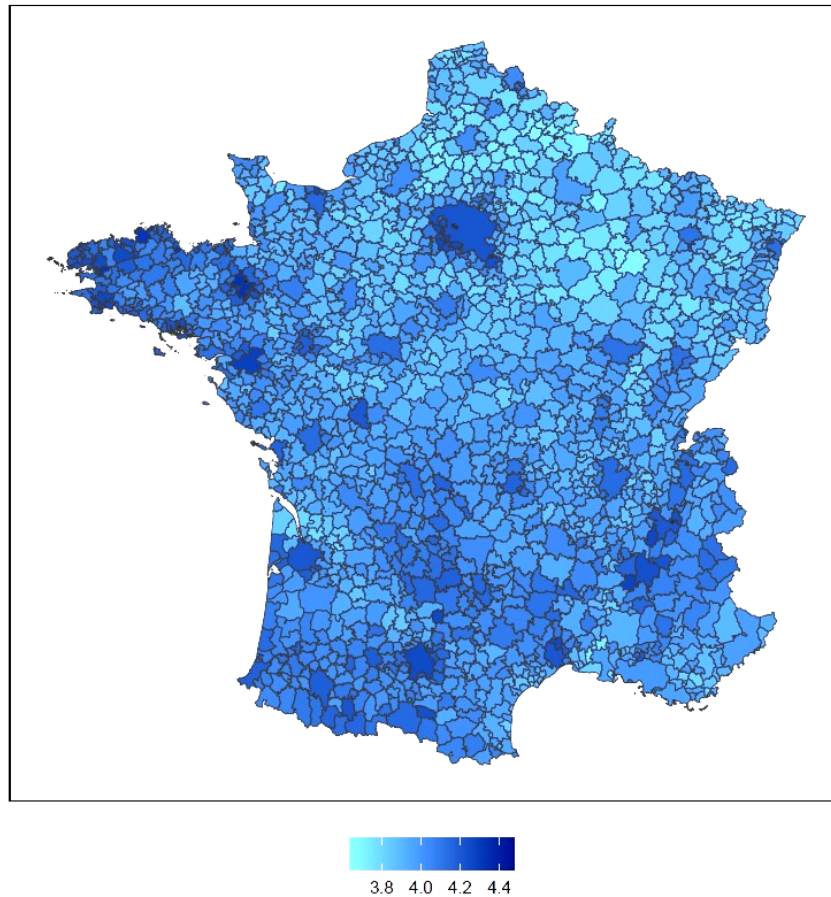


Figure 1 – Map of the local interpersonal trust variable

Several results stand out. First, the largest urban areas have the higher average level of trust, in particular Paris, Rennes, Nantes, Angers, Bordeaux, Toulouse, Montpellier, Strasbourg, Lille, Lyon and Grenoble, with a few exceptions such as Marseille, Nice, Toulon, and Perpignan in the South, as well as Douai, Le Havre, Lens in the North which do not stand out on this map. On the other hand, rural areas are not all characterized by low levels of interpersonal trust : while Brittany, the South-West, the Alps and the South of Massif Central are characterized by high levels of trust, the North-East, the Côte d'Azur, the region north of Bordeaux are characterized by low levels of interpersonal trust.

The distribution of the local synthetic interpersonal trust variable provides in effect a wholly new metric that synthesizes several key socio-demographic and political local determinants. No single variable alone can account for the observed distribution in the synthetic local interpersonal trust variable, be it 2017 electoral results or socio-demographic variables.

## 2.d Local interpersonal trust and participation to the *Grand Débat National*

### 2.d.1 Participation variables of interest

I consider various metrics of participation to GDN, both online and offline, at the intensive and extensive margins. All of these variables are aggregated at the life zone level, through sums and averages.

To measure offline participation, two variables are considered :

- Local meetings : number of local meetings reported on the GDN website, per 100 000 inhabitants
- Attendance : average number of participants reported for these meetings (this variable is not available for 236 life zones where either no local meeting was organized, or these meetings did not report attendance figures).

The former variable can be considered to measure the extensive margin of offline participation, while the latter to measure the intensive margin.

To measure online participation, five variables are considered :

- Online contributors : number of unique contributors to the online platform per 100 000 inhabitants.
- Close-ended questions : average number of closed-ended questions answered by these contributors
- Open-ended questions : average number of open-ended questions answered by these contributors
- Total word count : average total number of words used by these contributors in the open-ended questions
- Words per open-ended question : average number of words used by these contributors per open-ended question

The first of these variables measures the extensive margin of online participation, while the four other variables measure the intensive margin of online participation.

Summary statistics of these variables are given in table 4. Most of these variables do not follow a clean geographical pattern that is worth commenting upon, except for the number of online contributors, represented in figure 2. This figure shows that participation is higher in large cities. On top of that, participation is higher in the South, especially the South East, and on the Atlantic coast, than in the rest of the country.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Local meetings	1,632	10.233	9.217	0.000	3.368	8.450	14.216	46.526
Attendance	1,396	44.266	26.406	4.866	26.000	38.779	54.689	150.000
Online contributors	1,632	521.384	195.162	101.434	384.406	483.743	625.381	1,126.786
Close-ended questions	1,632	23.563	1.545	16.893	22.687	23.628	24.443	29.233
Open-ended questions	1,632	9.698	1.890	2.625	8.555	9.674	10.836	18.086
Total word count	1,632	527.254	109.205	174.000	460.288	526.319	583.043	876.323
Words per open-ended question	1,632	55.419	11.892	21.259	48.192	54.172	60.354	103.578

Table 4 – Summary statistics of participation variables of interest



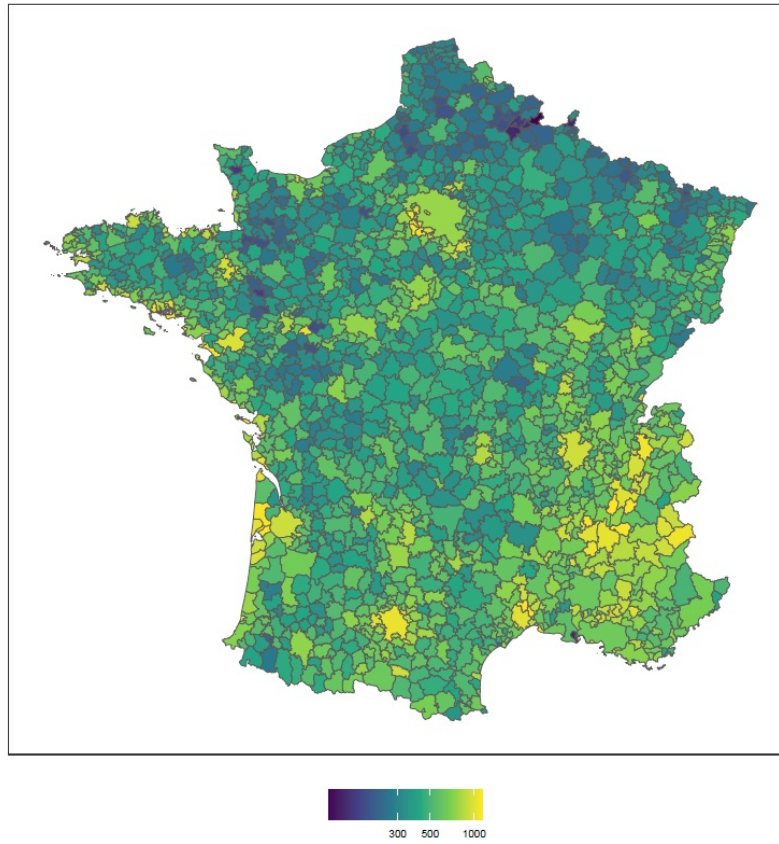


Figure 2 – Map of the density of online contributors (per 100 000 inhabitants)

	<i>Dependent variable:</i>				
	Online contributors (1)	Close-ended questions (2)	Open-ended questions (3)	Total word count (4)	Words per open-ended question (5)
Interpersonal trust	1,167.510*** (30.422)	0.270 (0.450)	3.189*** (0.534)	134.195*** (31.545)	-4.675 (3.463)
<i>Département</i> fixed effects and log-density as controls	X	X	X	X	X
Observations	1,632	1,632	1,632	1,632	1,632
R <sup>2</sup>	0.735	0.074	0.129	0.089	0.074

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 5 – Cross-sectional relationship between interpersonal trust and online participation

	<i>Dependent variable:</i>	
	Local meetings	Attendance
	(1)	(2)
Interpersonal trust	21.599*** (2.548)	1.983 (8.224)
<i>Département</i> fixed effects and log-density as controls	X	X
Observations	1,632	1,396
R <sup>2</sup>	0.165	0.130
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Table 6 – Cross-sectional relationship between interpersonal trust and offline participation

### 2.d.2 Cross-sectional results

Geographic cross-sectional regression results are given in tables 5 and 6. Appendix section 5.d.1 presents conditional scatter plots of these cross-section relationships.

Local interpersonal trust drives participation to GDN along certain dimensions, but not others. First of all, it drives online participation at the extensive margin. From table 5, more trusting life zones also have more online participants to GDN. The effect is significant and sizable : an increase of one standard deviation in the interpersonal trust variable (0.124) is associated with 145 more online contributors per 100 000 inhabitants. This result holds when looking at the extensive margin of offline participation : from table 6, more local meetings are organized in high-trust life zones. This finding echoes [Fourniau et al. \(2019\)](#) who find that more local GDN meetings take place in places with higher Macron vote share in 2017 (local interpersonal trust is indeed driven up by a higher Macron share of the vote). This supports the idea that interpersonal trust drives participation to GDN, and that GDN tended to attract participants with higher levels of interpersonal trust on average, while individuals with lower levels of interpersonal trust, such as the *Gilets Jaunes*, were relatively absent from it.

When it comes to the role played by interpersonal trust to explain the intensive margin of participation, the results are mixed. From table 6, interpersonal trust does not play a role in explaining the attendance of local meetings : more local meetings were organized in high-trust life zones, but these meetings do not attract more participants. Therefore, interpersonal trust does not drive offline participation at the intensive margin.

On the other hand, table 5 shows that higher local interpersonal trust is associated with more open-ended questions answered, which translates into a higher overall number of words used. Note that the higher overall number of words used does not come from longer answers to each question. This result implies that, on top of having contributed in higher numbers to the online platform, respondents in high-trust life zones have also answered more open-ended questions, which further amplifies their overrepresentation among all online contributions. Conversely, there were already fewer respondents from low-trust life zones, and they answered less open-ended questions, as a result, the relative weight of their answers in the overall mass of contributions is further diminished.

Finally, it appears that close-ended questions are not sensitive to interpersonal trust. Therefore, the answers given to close-ended questions are less prone to a selection based on interpersonal trust, relative to open-ended questions. It might be because close-ended questions are a less demanding and more accessible form of participation, while open-ended questions demand some language and expression skills which tend to be possessed more by high-trust individuals. Experimental evidence from section 3 will confirm this finding.

## 2.e Local interpersonal trust and lexical aspects of online contributions

### 2.e.1 The dictionary-based approach with the LIWC software

I analyse the content of answers to the open-ended questions in the contributions to GDN with a dictionary-based approach, which consists in considering the extent to which the analyzed texts match some predefined dictionaries ; said otherwise, whether certain words are present or not.

Due to their simplicity, dictionary-based approaches are popular text analysis methods. I opt for the dictionary-based software Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001; Tausczik and Pennebaker, 2010). This key contribution of this software consists in its rich set of built-in dictionaries that enable to analyze input texts across 80 categories : these dictionaries have been constructed and validated to be most psychologically meaningful. LIWC has already been mobilized to study economic questions, such as the changes in chair of the US Federal Reserve Alan Greenspan's language use across the economic cycle (Abe, 2011), linguistic features explaining success in peer to peer lending (Larrimore et al., 2011), or reviews on Airbnb (Quattrone et al., 2018).

I use the French translation of these dictionaries from Piolat et al. (2011) and analyse jointly all the answers that a respondent has given to the open-ended questions (see appendix section 5.a for the list of these questions). The answers of 197517 respondents are analyzed, i.e. all respondents who answered to at least one open-ended question in the four long questionnaires.

### 2.e.2 Lexical variables of interest

Out of the numerous variables generated by this software, I focus on these 9 variables, which are most salient for the analysis :

- Positive emotions (love, nice, sweet)
- Anxiety (worried, fearful)
- Sadness (crying, grief, sad)
- Anger (hate, kill, annoyed)
- Swear words (fuck, damn, shit)
- Negations (no, not, never)
- Exclamation marks
- 6 letter words
- Word syllables
- Sentence length

The six first variables are dictionary-based lexicon variables that correspond to the share (in percent) of words matching a predefined dictionary. Examples of the words used in each dictionary are given between parentheses. The variables "Anxiety", "Sadness" and "Anger" correspond to the breakdown a "negative emotions" variable, while no such breakdown exists for the positive emotions variable. Negations is a dictionary-based grammatical variable measuring the share of words that are used to express a negation within a given text. Exclamation marks variable measures the occurrence of exclamation marks in a given text.

On top of these variables, lexical complexity of the texts is measured with 3 variables : the share of words above 6 letters, the average number of syllables by word, and the average number of words per sentence. These variables are used extensively in English linguistics to measure the complexity (also commonly referred to as readability) of texts, and to build indices thereof. Due to the absence of consensual equivalent indices for the French language (Bossé-Andrieu, 1993), I opt to study these variables separately and not aggregated into an index.

After having being computed at the individual level, for each individual contribution, these variables are averaged at the life zone level, using the reported zipcode given by the respondent, discarding invalid zipcodes, as well as respondents that do not self-describe as “Citizens”. 166683 respondents remain. Summary statistics of these variables across life zones are given in table 7.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Positive emotions	1,632	2.919	0.296	2.078	2.747	2.932	3.092	3.780
Anxiety	1,632	0.104	0.035	0.016	0.083	0.104	0.122	0.210
Sadness	1,632	0.384	0.078	0.183	0.336	0.383	0.428	0.622
Anger	1,632	0.283	0.071	0.112	0.241	0.279	0.321	0.495
Swear words	1,632	0.015	0.011	0.000	0.007	0.014	0.020	0.052
Negations	1,632	2.507	0.331	1.658	2.313	2.482	2.684	3.574
Exclamation marks	1,632	0.523	0.187	0.107	0.407	0.502	0.621	1.128
6 letter words	1,632	31.119	0.991	28.285	30.585	31.180	31.732	33.623
Word syllables	1,632	1.804	0.030	1.725	1.785	1.804	1.821	1.885
Sentence length	1,632	17.691	1.739	13.432	16.640	17.654	18.615	23.117

Table 7 – Summary statistics of lexical variables of interest

### 2.e.3 Cross-sectional results

Geographic cross-sectional regression results are given in tables 8 and 9. Appendix section 5.d.2 presents conditional scatter plots of these cross-section relationships.

From table 8, respondents in high interpersonal trust life zones tend to use a more anxious and sad lexicon, while respondents in low interpersonal trust life zones tend to use a more angry lexicon, with more negations and exclamation marks. However, no relationship is uncovered with swear words and positive emotions lexicon.

These results is quite telling : respondents in low interpersonal trust life zones used GDN to channel more anger, virulence, and negativity, however, it did not translate into mere vulgarity. This result can reflect the fact that low-interpersonal trust translates into higher *Gilets Jaunes* identification and support (Algan et al., 2019b). On the other hand, respondents in life zones with a higher level of interpersonal trust, perhaps surprisingly, did not use the GDN to express more positive emotions, mirroring respondents in low-trust life zones. Instead, they expressed more anxiety and sadness, which may reflect the tense political climate of the time : these respondents did not share the anger of *Gilets Jaunes*, but still they expressed other forms of negative emotions in their answers.

Interpersonal trust, according to table 9, is positively correlated to the three measures of lexical complexity. The fact that a positive effect shows up for all three variables adds confidence in the robustness of this result : respondents in high-trust life zones use more complex language to answer to the open-ended questions. Apart from reflecting the well-established social disparities in language, the disparities in lexical complexity for different levels of interpersonal trust could be relevant when it comes to the capacity to influence policies : it could translate into a differential treatment of these contributions, if for instance the algorithm studying contributions to GDN is more apt at analysing simple texts than more complex ones (or the other way around).

In short, these results point at the lexical diversity of online contributions depending on the level of interpersonal trust. The level of interpersonal trust does not only lead to difference in participation to GDN, but also to the tone used in the online contributions.

	<i>Dependent variable:</i>						
	Positive emotions	Anxiety	Sadness	Anger	Swear words	Negations	Exclamation marks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interpersonal trust	-0.104 (0.087)	0.019* (0.010)	0.063*** (0.023)	-0.043** (0.021)	0.003 (0.003)	-0.396*** (0.095)	-0.195*** (0.054)
<i>Département</i> fixed effects and log-density as controls	X	X	X	X	X	X	X
Observations	1,632	1,632	1,632	1,632	1,632	1,632	1,632
R <sup>2</sup>	0.064	0.064	0.072	0.085	0.086	0.101	0.100

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 8 – Cross-sectional relationship between local interpersonal trust and dictionary-based lexical variables of interest

	<i>Dependent variable:</i>		
	Share of six-letter words	Mean word syllables	Mean sentence length
	(1)	(2)	(3)
Interpersonal trust	1.821*** (0.278)	0.041*** (0.009)	1.218** (0.503)
<i>Département</i> fixed effects and log-density as controls	X	X	X
Observations	1,632	1,632	1,632
R <sup>2</sup>	0.140	0.095	0.088

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 9 – Cross-sectional relationship between local interpersonal trust and lexical complexity variables

## 2.f Local interpersonal trust and profiles of online contributors

### 2.f.1 Construction of the profiles of respondents to the short questionnaires

I build profiles of respondents by applying the Latent Dirichlet Allocation (LDA) algorithm to answers to the GDN. Originally, this machine learning algorithm is a probabilistic topic model that performs as an unsupervised clustering tool. It has been used mainly in the artificial intelligence literature on natural language analysis to generate topics from a set of texts (Blei et al., 2003). This powerful tool calls for applications in text analysis for social sciences (Gentzkow et al., 2019). It has been used in Draca and Schwarz (2018) and Stantcheva (2019), in a novel way, to cluster answers to survey questionnaires into political profiles of respondents that hold strikingly distinct positions on the different topics. The paramount contribution of using LDA is that these profiles are created in a bottom-up, unsupervised way. No priors are imposed about the construction of these profiles.

I use their method to build four profiles of respondents to the short questionnaires, and assign each respondent to one profile. Appendix section 5.c.1 describes in detail the way in which the LDA algorithm performs the creation of profiles and the assignment of each respondent into one profile.

To build these four profiles of respondents, I use answers to the close-ended questions from the short questionnaires. I consider 334636 respondents who answered to at least one question to the short questionnaire.

Using the top answers per profile (see appendix section 5.c.3), which are the most salient answers for a given profile, I characterize each profile in the following way :

- Profile 1 : “Gilet Jaune sympathizer” : oppose taxes on fuel, support referenda at a local and national levels, display hostility to the State and stronger attachment to devolved authorities, and express demands for more in-person help when dealing with administrative issues.
- Profile 2 : “Environmentalism” : support any action to protect the environment, including new taxes.
- Profile 3 : “Supporter of democratic reforms” : support institutional reforms for more direct democracy, proportional representation, as well as compulsory voting. Also support some improvements in how the state administration works.
- Profile 4 : “Small government proponent” : support reforms to make state administration more efficient, although they are not themselves facing difficulties with the administration, champion the decrease in public spending as well as the reduction in the number of parliamentarians, and oppose institutional reforms such as referenda and proportional representation.

Each respondent to the short questionnaire is then assigned to one of these profiles. Table 10 shows the distribution of respondents into the four profiles. Overall, the assignment creates four groups of roughly equal size, even if more respondents are assigned to profiles 1 and 2 than to profiles 3 and 4.

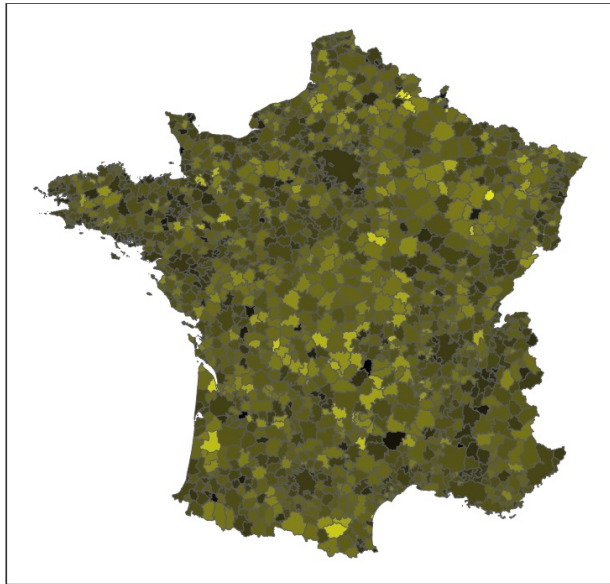
LDA profile	Number of respondents	Share of all respondents (in percent)
1	98510	29,4
2	91920	27,5
3	72968	21,8
4	71238	21,3
Total	334636	100

Table 10 – Distribution of respondents into the 4 LDA profiles (short questionnaire)

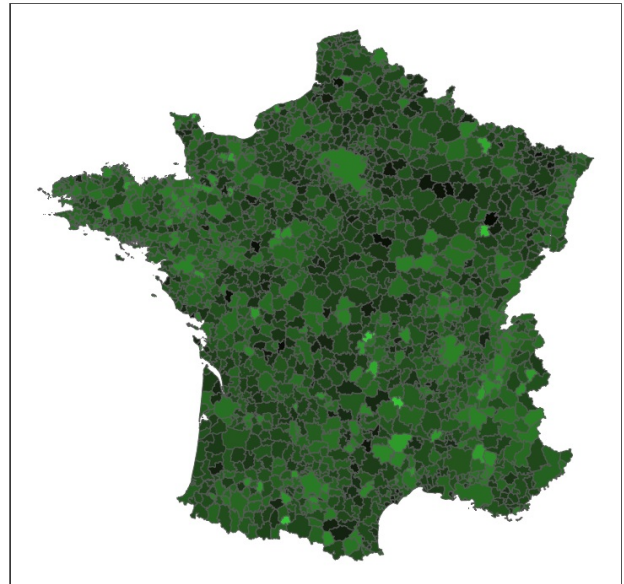
Respondents are then aggregated per profile at the life zone level, and the local shares of each profile at the life zone level are computed. Summary statistics of these variables are given in table 11. The geographic distribution of these variables is mapped out in figure 3. Profile 1 respondents are concentrated in the notorious “diagonal of low densities” from Ardennes in the North East to Landes in the South West, while profile 2 respondents tend to concentrate everywhere else. No such geographic pattern stands out for the distribution of profile 3 and 4 respondents.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Share of profile 1 respondents	1,632	33.585	7.321	10.702	28.866	33.113	37.805	70.591
Share of profile 2 respondents	1,632	22.720	6.140	3.030	18.718	22.424	26.538	49.250
Share of profile 3 respondents	1,632	22.993	5.599	0.000	19.859	22.837	26.118	50.000
Share of profile 4 respondents	1,632	20.702	5.295	0.000	17.496	20.589	23.713	50.317

Table 11 – Summary statistics of the local share of respondents assigned to each short questionnaire profile



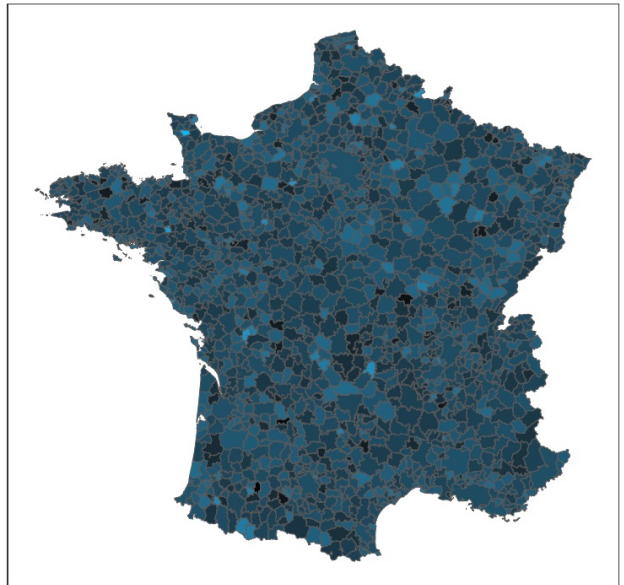
(a) Profile 1 - Gilets Jaunes sympathizer



(b) Profile 2 - Environmentalist



(c) Profile 3 - Supporter of democratic reforms



(d) Profile 4 - Small government proponent

Figure 3 – Maps of the local shares of respondents assigned to each short questionnaire profiles



## 2.f.2 Construction of the profiles of respondents to the “Democracy and Citizenship” long questionnaire

Two profiles of respondents are created using the answers to the “Democracy and Citizenship” long questionnaire, and applying a novel approach based on LDA, described in detail in appendix section 5.c.2. This novel approach builds on [Draca and Schwarz \(2018\)](#) and [Stantcheva \(2019\)](#), it combines aspects of supervised (keyword approach) and unsupervised text analysis, and enables to build profiles of respondents in a flexible manner, while sticking as much as possible to the characteristics of the data that we are considering (i.e. a combination of close-ended and open-ended questions).

Following the described methodology, the answers of 80598<sup>3</sup> respondents to the “Democracy and Citizenship” long questionnaire are analyzed with LDA, and the respondents are sorted into 2 profiles.

The top answers per profile are given in appendix section 5.c.4 and enable to characterize the two profiles as follows :

- Profile 1 : “Right-wing respondent”, who favors strict immigration policies, counterparts for people taking social benefits, the reduction in the number of elected representatives, and sanctions to deal with incivilities.
- Profile 2 : “Left-wing respondent”, who favors a stronger role for unions and associations, more solidarity, no counterpart for people taking social benefits, no tightening of immigration laws.

Summary statistics for the number of respondents to the “Democracy and Citizenship” questionnaire classified into either profiles are given in table 12. The algorithm has sorted respondents to this questionnaire into two groups that are roughly similar in size.

LDA profile	Number of respondents	Share of all respondents (in percent)
1	42959	53,3
2	37639	46,7
Total	80598	100

Table 12 – Distribution of respondents into the two profiles for the “Democracy and Citizenship” long questionnaire

Respondents are then aggregated per profile at the life zone level<sup>4</sup>. Summary statistics of the local share of respondents that belong to profile 1 are given in table 13<sup>5</sup>, while figure 4 maps out the life zones where respondents assigned to profile 1 (resp. profile 2) outnumber that of the other profile. The distribution of the profiles of respondents seem to be in line with the historical political leaning of the region : indeed, this map is reminiscent of electoral maps such as the second round of the 2007 and 2012 French presidential elections, with a divide between left-wing regions in the West, South-West and Central regions of France, and right-wing regions in the East and Côte d’Azur ([Le Bras and Todd, 2013](#)).

<sup>3</sup>99229 respondents answered to this questionnaire, however, some respondents could not be taken into consideration because they did not answer to any close-ended question and did not mention any of the predefined keywords in their answers.

<sup>4</sup>For one life zone where no respondent answered to the “Democracy and citizenship” long questionnaire, the share of respondents belonging to either profile cannot be computed.

<sup>5</sup>Only the share of profile 1 respondents is given, since respondents are uniquely assigned to either profile 1 or profile 2.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Share of profile 1 respondents	1,631	51.951	14.341	0.000	44.642	60.085	100.000

Table 13 – Summary statistics of the local share of respondents assigned to the “Democracy and citizenship” long questionnaire profiles

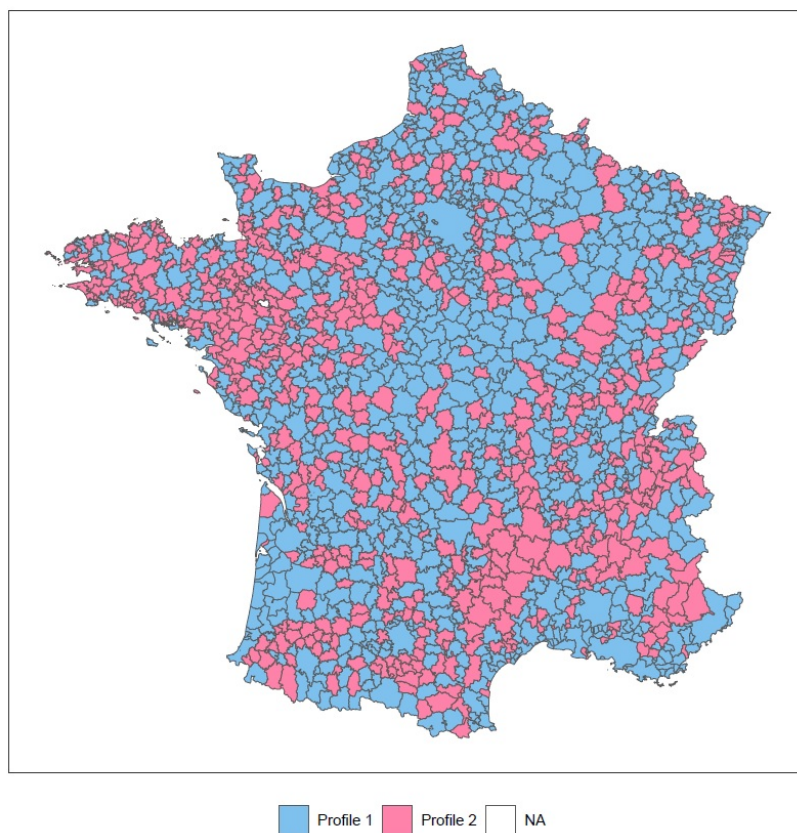


Figure 4 – Map of the life zones where each profile is more numerous (LDA profiles for the “Democracy and Citizenship” long questionnaire)

### 2.f.3 Cross-sectional results

Geographic cross-sectional regression results are given in tables 14 and 15. Appendix section 5.d.3 presents conditional scatter plots of these cross-section relationships.

From table 14, local interpersonal trust is associated to certain profiles of respondents to the short questionnaires and less to others. In particular, more trusting life zones have more environmentalist contributors, whereas less trusting life zones have more *Gilets Jaunes* sympathetic contributors, a result that is consistent with [Algan et al. \(2019b\)](#). A weaker correlation exists with the two other profiles : higher local interpersonal trust is associated with less respondents demanding democratic reforms and with more respondents who are in favor of small government. This result implies that the content of online contributions reflects the local level of interpersonal trust, and hints at the existence of a gap in the content expressed by low-trust individuals compared to high-trust individuals. The latter had more to say about environmental protection and demands for a smaller government, while the former tended to use the GDN online platform to express demands about democratic and administrative reforms, as well as the repeal of the taxes on fuel. Since a higher level of interpersonal trust is associated with greater participation, this result implies that the demands for higher environmental protection and smaller government tended to be overrepresented compared to the demands for democratic and administrative reforms, among all online contributions. This finding should be taken into account when interpreting the results of the GDN.

However, from table 15, I fail to detect any correlation between interpersonal trust and the shares of respondents belonging to either profile for the “Democracy and citizenship” long questionnaire. The left-right divide identified among respondents to this questionnaire does not correlate with interpersonal trust. This indicates either that our geographical cross-section approach is not precise enough to detect any effect in this instance, or that the traditional left-right divide is not driven primarily by interpersonal trust, a result reminiscent of [Algan et al. \(2018\)](#).

	<i>Dependent variable:</i>			
	Profile 1	Profile 2	Profile 3	Profile 4
Interpersonal trust	-24.719*** (1.928)	24.201*** (1.564)	-5.782*** (1.621)	6.300*** (1.540)
<i>Département</i> fixed effects and log-density as controls	X	X	X	X
Observations	1,632	1,632	1,632	1,632
R <sup>2</sup>	0.243	0.292	0.085	0.076

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 14 – Cross-sectional relationship between local interpersonal trust and the profiles of respondents for the short questionnaires

	<i>Dependent variable:</i>
	Profile 1
Interpersonal trust	3.294 (4.121)
<i>Département</i> fixed effects and log-density as controls	X
Observations	1,631
R <sup>2</sup>	0.099
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 15 – Cross-sectional relationship between local interpersonal trust and the profiles of respondents for the long “Democracy and citizenship” questionnaire

## 2.g Robustness checks

The cross-sectional results are robust to alternative measures of local interpersonal trust, such as the synthetic local trust variable from [Algan et al. \(2019b\)](#).

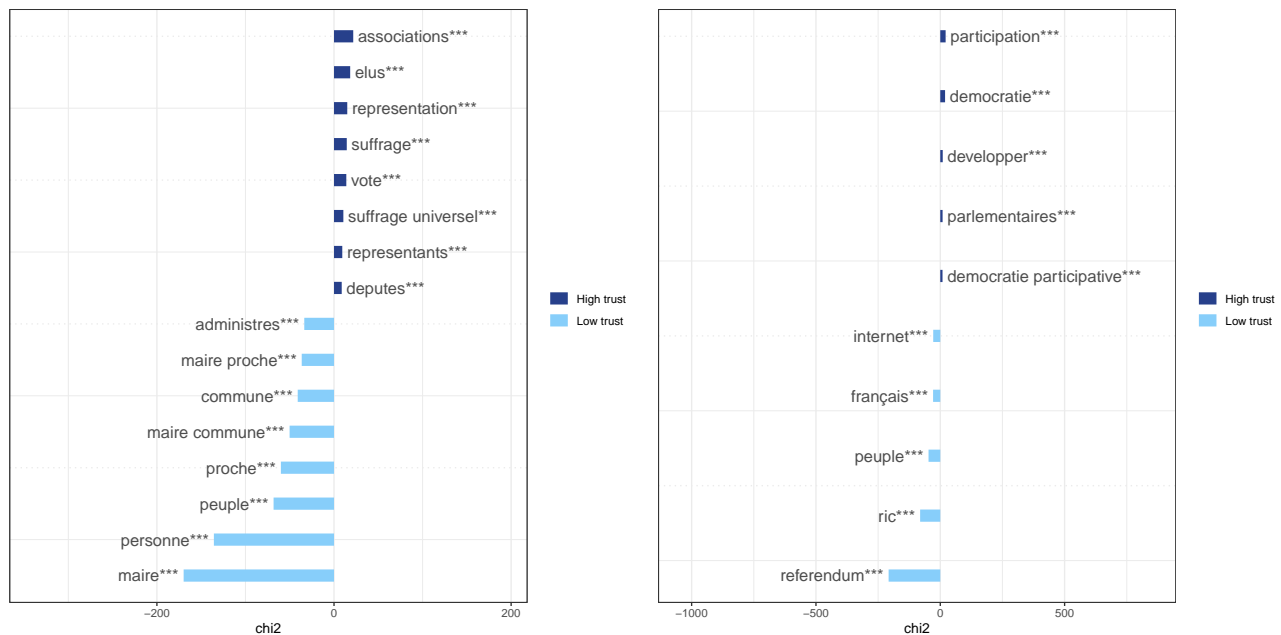
On top of that, similar regressions are run, where the synthetic trust variable is broken into the various local demographic, social and political variables that are used to build it (see appendix section 5.e). This aims at showing that no single socio-demographic or political variable is driving all the previous regression results. What this shows is that the synthetic local interpersonal trust variable is indeed a novel aggregated metric capturing effects that are otherwise not visible with a regression on disaggregated variables.

## 2.h Keynes analysis : local interpersonal trust and lexical content of online contributions

Keyness graphs are a useful tool to represent the textual diversity of texts across a dichotomous variable. This approach has been carried out in [Stantcheva \(2019\)](#) and is equivalent to the one used in [Gentzkow and Shapiro \(2010\)](#). The production of keyness graphs rests on the statistical package [Quanteda](#) by [Benoit et al. \(2018\)](#).

Keyness analysis rests on the comparison of the frequency of appearance of certain phrases<sup>6</sup> in a given group compared to another group - in this case, respondents from life zones with a low level of interpersonal trust (below the median level of local interpersonal trust) compared to respondents from life zones with a high level of interpersonal trust (above the median). Appendix figure 22 maps out the geographic distribution of low-trust and high-trust life zones as defined here.

This approach drastically reduces the textual complexity of the answers by reducing texts to a list of words or groups of words (see appendix 5.f.2 for details on the text preprocessing). As such, it disregards the meaning of texts. However, it performs well to get a glimpse at the lexical diversity of answers across life zones for different levels of trust. Appendix section 5.f.3 explains the construction of the keyness statistic used to build the keyness graphs.



(a) *En qui faites-vous le plus confiance pour vous faire représenter dans la société et pourquoi ?*  
 Who do you trust more to represent you in society and why?  
 (b) *Que faudrait-il faire aujourd’hui pour mieux associer les citoyens aux grandes orientations et à la décision publique ?*  
 What should be done today to better involve citizens in major orientations and in public decision-making?

Figure 5 – Keyness graphs representing textual diversity of answers in low-trust relative to high-trust life zones

<sup>6</sup>“Phrases” here means words or combination of words

Not all questions could be analyzed with keyness analysis. Whenever a question was asking for a complex reasoning or some element of personal experience, the reduction of answers to a list of words or groups of words is a hindrance. However, for more simple questions, keyness analysis is arguably more telling. I decided to focus on two questions for which the diversity in answers, as highlighted by the keyness analysis, was best apparent across high-trust and low-trust life zones. The results for these two questions are also most relevant to our research question on interpersonal trust and participatory democracy.

The question “Who do you trust more to represent you in society and why?” shows a very stark contrast in answers between the words and expressions that appear relatively more in low-trust life zones compared to high-trust life zones. In low-trust life zones, respondents tend to mention more the mayor (“maire”), on grounds that he/she is closer (“proche”) to the citizens, or they answer that they trust nobody (“personne”). Note that the latter result provides additional evidence for the validity of our synthetic local interpersonal trust variable. On the other hand, respondents in high trust life zones mention elected representatives through universal suffrage (“suffrage universel”, “vote”, “représentation”, “élus”, “députés”), as well as associations. The fact that respondents mention associations more in high-trust life zones echoes the findings of Putnam (2000) : places where social capital is higher tend to have more associations (see appendix figure 13) and citizens display more interpersonal trust. The contrast between low-trust and high-trust life zones in their trust for elected officials is striking and reflects a divide between some regions where individuals are more defiant of politicians in general, except for the mayor, which are regions characterized by low levels of interpersonal trust, and other regions where elected officials are less distrusted, which are regions with a higher level of interpersonal trust.

As for the question “What should be done today to better involve citizens in major orientations and in public decision-making?”, in low-trust life zones, respondents mention the referendum (“referendum”, “ric”, which stands for *référendum d’initiative citoyenne*, citizen’s initiative referendum, a key demand of participants to the *Gilets Jaunes* movement) relatively more than respondents in high-trust life zones, who tend to mention participatory democracy more (“démocratie participative”). This finding complements those of section 2.d on the predilection for participatory democracy in more trusting life zones : respondents in these life zones not only participate more in participatory democracy such as GDN, but also mention it relatively more as an attractive outlook for all citizens. Respondents in low-trust life zones, on the other hand, while taking part in participatory democratic tools such as GDN, favor more direct democratic ways to involve citizens into public decision-making, such as the referendum.

All in all, keyness analysis highlights the disparities between low-trust and high-trust life zones at the textual level : the words used by respondents are different, hinting at a gap in the demands expressed by individuals depending on their level of interpersonal trust. This finding runs counter the idea of a uniform population that took part in GDN, uniform in its demands and aspirations. In fact, there are some geographical disparities reflecting disparities within the participants themselves for different levels of interpersonal trust.

### 3 Experimental evidence on interpersonal trust and participation to public consultations

The previous section has shown that, when using actual GDN data and a geographic cross-sectional approach, a synthetic local interpersonal trust variable correlates with various measures of GDN participation. In particular, more trusting places participate more to GDN, both at the extensive and intensive margins, and the content, as well as lexical and textual aspects of their contributions, differ from that of less trusting places.

To strengthen our claim that interpersonal trust is indeed a key factor driving participation to public consultations in general, I conducted an pilot experiment, which, despite its limited scale, provides additional evidence on the decisive role of interpersonal trust in explaining participation to public consultations, both at the extensive and intensive margins. The experiment concludes that subjects with a higher level of interpersonal trust will tend to answer to more questions, have longer answers and spend more time on the experimental questionnaire than less trusting subjects. This experiment thus confirms the claims made in section 2.d. However, due to the experimental design and limited sample size, the experiment cannot confirm the results from sections 2.e, 2.f, and 2.h, regarding lexical, textual and content-wise aspects of the contributions to public consultations.

#### 3.a Purpose of the experiment and related literature

This experiment aims at testing the hypothesis that interpersonal trust drives voluntary participation to a questionnaire akin to GDN questionnaires. For the purpose of this experiment, I depart from the measurement of trust that is mobilized above and which makes use of the traditional question on trust (“Generally speaking, would you say that most people can be trusted, or that you can never be too careful when dealing with others ?”). Instead, I build on [Glaeser et al. \(2000\)](#), which uses experimental designs in order to measure trust, and rely on the trust game put forward in [Berg et al. \(1995\)](#) Although [Glaeser et al. \(2000\)](#) propose a slightly modified version of the original trust game, this experiment relies on the original version. This is for practical reasons : the original version of the trust game can be implemented easily with online individual experiments like here, whereas the version in [Glaeser et al. \(2000\)](#) requires communication between players, hence the use of laboratories.

The trust game from [Berg et al. \(1995\)](#) has been used extensively to measure trust in the literature : [Fershtman and Gneezy \(2001\)](#) use it to study trust between different ethnic groups in Israel, [Carlin and Love \(2013\)](#) to study partisanship, and [Willinger et al. \(2003\)](#) the intercultural differences between France and Germany. This empirical design has been so popular, that [Johnson and Mislin \(2011\)](#) reports 162 replications of the game. However, note that the trust game from [Berg et al. \(1995\)](#) has been criticized on the grounds that it measures more of individual trustworthiness than actual trust ([Glaeser et al., 2000](#)), though here, I rely on the standard interpretation that posits that the trust game measures interpersonal trust.

#### 3.b An experiment on interpersonal trust and participation to an online public consultation

##### 3.b.1 Experimental design

The experiment has been built using the software Qualtrics<sup>7</sup>. It has two parts. First, a game is played to measure and manipulate the subject’s trust, mobilizing the trust game put forward in [Berg et al. \(1995\)](#).

---

<sup>7</sup>It can be accessed at the following link : <https://harvard.az1.qualtrics.com/jfe/form/SV9G0m7sbvZkBnrzD>.

Then, subjects are invited to answer a questionnaire which is designed to look like public consultations such as GDN.

Screenshots of all the instructions can be found in section 5.g.2 of the appendix. The full questionnaire of the second part can be found in section 5.g.3 of the appendix.

In the first part, the subject plays the trust game designed in [Berg et al. \(1995\)](#). In this game, two players, the trustor and the trustee, are both given 10 (experimental) points at the beginning of the game. Then the trustor is given the possibility to send some of these points (integer amount) to the trustee. These points are tripled and given to the trustee. The trustee finally has the possibility to reciprocate and send points back to the trustor. The amount sent by the trustor measures this person's level of trust towards the other person. Conversely, the amount which is sent back by the trustee is interpreted as the extent of the trustee's trustworthiness.

In this experiment, the subjects play the game as trustors. Data points randomly drawn from [Berg et al. \(1995\)](#) (no social history treatment) are used to determine the amount which is sent back by the trustee. When for example the subject sends 7 points, a data point (one of the trustees) is drawn at random from [Berg et al. \(1995\)](#), among the observations where the trustor sent 7, to determine which amount is sent back. In this example, there are 3 such data points in the [Berg et al. \(1995\)](#) experiment, two where 1 point was sent back, and one where 6 points were sent back. The subject who sent 7 points has thus one in three chances of receiving 6 points back, and two in three chances of receiving 1 point back. The same randomization occurs for all amounts sent by the subject to this experiment<sup>8</sup>.

In terms of incentives (to make subjects play this game seriously), players are payed in "experimental points" which are lottery tickets : the more points they get, the higher their chance of winning a prize in a lottery carried out once the experiment is completed. This design helps circumvent a possible issue arising due to risk aversion, as pointed out by [Roth \(1986\)](#).

This first section serves two purposes. First, building on [Glaeser et al. \(2000\)](#), it elicits the subject's level of interpersonal trust from the amount the person sends in the trust game as trustor. This higher the amount sent, the more trusting the subject is revealed to be. Indeed, in a game where the Nash equilibrium is to not send anything, sending any amount to the other person reveals that the subject trusts the other individual for sending back at least this amount at the second stage of the game. Instead of asking a simple survey question on trust ("do you trust X?") which has no actual consequences, this game experimentally reveals the subject's level of interpersonal trust, in a practical setting, with real stakes. Note that subjects are told explicitly that they are playing against a stranger whose answers have been recorded in advance. Therefore, the amount they send should be interpreted as revealing their trust in strangers, i.e. interpersonal (or general) trust.

Furthermore, this game is also used to experimentally manipulate the subject's level of trust in others, through a random treatment. This happens via the random amount which is sent back by the other player. Everything else held constant, the higher the amount which is sent back, the higher the subject's (manipulated) level of interpersonal trust. Indeed, the higher the amount which is sent back, by a stranger who had no incentive to send back any points, the more the subject will tend to believe that other individuals are trustworthy. This is a random treatment, since the amount which is sent back is picked randomly among the data points from [Berg et al. \(1995\)](#).

In the second section, the subject was offered the possibility to participate to a questionnaire on a topical issue. Due to the circumstances during which the experiment was run, instead of asking similar questions to the ones from GDN, the topic picked was the economic consequences of the COVID-19 crisis. Participation to this section was completely voluntary and the subject was told that the whole section could be

---

<sup>8</sup>With one exception : this experiment has data points for any amount sent, except 9, for which the data point from the social history treatment of [Berg et al. \(1995\)](#) is used (0 is sent back by the trustee).



skipped without any loss in payment. The questionnaire consisted in 3 close-ended questions first, and then 5 open-ended questions. At any time, the subject was offered the possibility to skip any question, or even to skip the whole section and directly validate the experiment. The questions were carefully thought of to be simple enough so that most subjects would have an opinion to give<sup>9</sup>, but challenging enough to keep them interested. The open-ended questions were designed to be potentially answered with great lengths.

An attention check was added right before the trust game (see appendix figure 27a). This attention check served three purposes. First, to disqualify subjects who did not pay enough attention and answer to surveys randomly<sup>10</sup>. It also mobilized the attention of the rest of the subjects for the actual trust game. Finally, it gave the rest of the subjects a preview of what the actual trust game question looks like, which helped clarify the game instructions.

### 3.b.2 Tested hypotheses

Participation to the questionnaire is measured with 5 variables :

- Number of close-ended questions answered (from 0 to 3)
- Number of open-ended questions answered (from 0 to 5)
- Total number of words used to answer the open-ended questions (total word count)
- Average number of words used per open-ended question answered (average word count)
- Total duration of the experiment, in seconds

These variables mirror the ones analyzed in the first section. Due to the limited scale of the experiment, further analysis of the content of the answers, with lexical and readability measures for instance, has proven inconclusive.

Two hypotheses are tested :

- The subject's initial level of interpersonal trust (as measured by the amount sent by the subject) is positively correlated with participation to the questionnaire.
- The subject's trust manipulation (as measured by the amount which is randomly sent back) correlates positively with participation to the questionnaire. This is akin to a random trust treatment, since the amount which is sent back is picked randomly among the data points from [Berg et al. \(1995\)](#), therefore the impact of trust manipulation on participation can be interpreted as causal.

### 3.b.3 Data collection

Data was collected via Amazon Mechanical Turk, in April and May 2020. 111 subjects took part in the experiment, 100 of them completed it, among which 27 failed the attention check. The final sample to be considered thus includes 73 subjects.

The experiment was conducted in French, and subjects had to be located in France to be able to access it. Individuals were paid between 1\$ and 2,5\$ as a fixed fee for taking part to the experiment : the pay was adjusted upwards during the course of the experiment in order to attract more subjects. Subjects who failed the attention check did not qualify for the fixed fee.

Subjects could also earn an extra 3\$ thanks to a lottery : their probability of winning, in percent, is equal to the number of points that they earned during the trust game section. The lottery and the payment of the 3\$ cash prize were conducted within the few days after the experiment was completed by a given subject.

---

<sup>9</sup>In fact, very few subjects answered "I don't know" or equivalent sentences.

<sup>10</sup>On Amazon Mechanical Turk, the quality of subjects and the effort put in answering surveys can be very diverse and is hard to control for with the built-in tools.

### 3.c Descriptive statistics

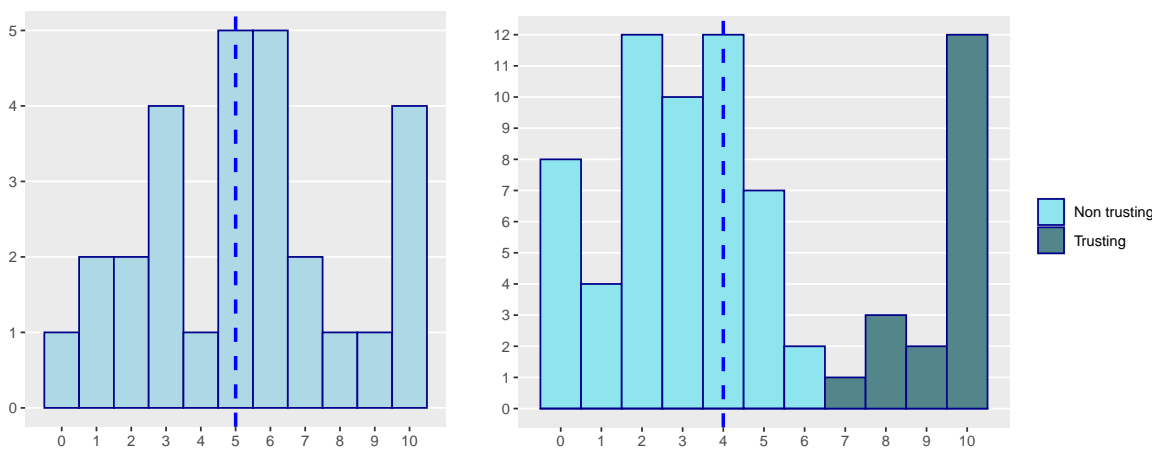
#### 3.c.1 Trust game variables

Overall, the distribution of amounts (experimental points) sent by subjects in the trust game is similar to the distribution in Berg et al. (1995) (under the no social history treatment) : see table 16 and figure 6. Although values are slightly lower in this experiment (median of 4 compared to 5, indicated by the dotted blue line), and are slightly more dispersed, most values in both experiments are concentrated around 5, and both distributions exhibit a peak at 10. A Wilcoxon rank-sum test fails to reject the null hypothesis of equal distributions across the two samples at standard levels (p-value = 0.1128).

Note that the distribution of amounts sent in our experiment follows a notable two-humps shape, with a first hump of values below 7 and and second hump above 7. For the subsequent regression analysis, I use the two humps of the distribution to conveniently divide the sample between “trusting” individuals, who sent 7 points or more in the trust game, and “non-trusting” individuals. Using this definition, our sample includes 18 trusting individuals (25% of the sample).

Statistic	N	Mean	St. Dev.	Min	Median	Max
Berg et al. (1995) experiment	28	5.286	2.904	0	5	10
This experiment	73	4.411	3.257	0	4	10

Table 16 – Summary statistics of the amount sent in both experiments



(a) Amounts sent in the trust game experiment by Berg et al. (1995) (no social history treatment)

(b) Amounts sent in this experiment

Figure 6 – Histograms of amounts sent in both trust games

The randomization of amounts sent back using data points (trustees) from the Berg et al. (1995) experiment leads to having a distribution of the pairs of amounts sent and received back that are very similar for both experiments (see appendix figure 23).

To summarize the second-stage of the trust game, I consider the number of points which are sent back as a share of the number of points sent in the first stage of the game. This variable is subsequently called “share sent back”. Indeed, the amount which is sent back is typically increasing with the amount sent (see

appendix figure 23), therefore I scale the amount sent back by the amount sent to get a variable which has the same order of magnitude whatever the amount sent by the subject in the trust game. For subjects who sent 0 and received 0 back, this variable is set to 100. Note that alternative solutions where the share sent back is coded as 0 or as missing have been considered, and lead to similar results. Summary statistics of this variable are given in table 17 and the histogram of this variable appears in figure 7.

In the subsequent regression analyses, I consider subjects who received more than twice the amount that they sent, as having being exposed to a “trust treatment” : for these subjects, the variable “share sent back” is 200 or above. This threshold corresponds to twice the value of the median. Indeed, for these subjects, the amount which is sent back by the other player in the trust game is far greater than the one that they sent, and indicates a great willingness from the other player to share gains from the game, i.e. to be trustworthy. Using this definition, our sample includes 17 subjects which were exposed to the “trust treatment” (23% of the sample).

Statistic	N	Mean	St. Dev.	Min	Median	Max
Share sent back	73	94.097	82.351	0	100	300

Table 17 – Summary statistics of the “Share sent back” variable

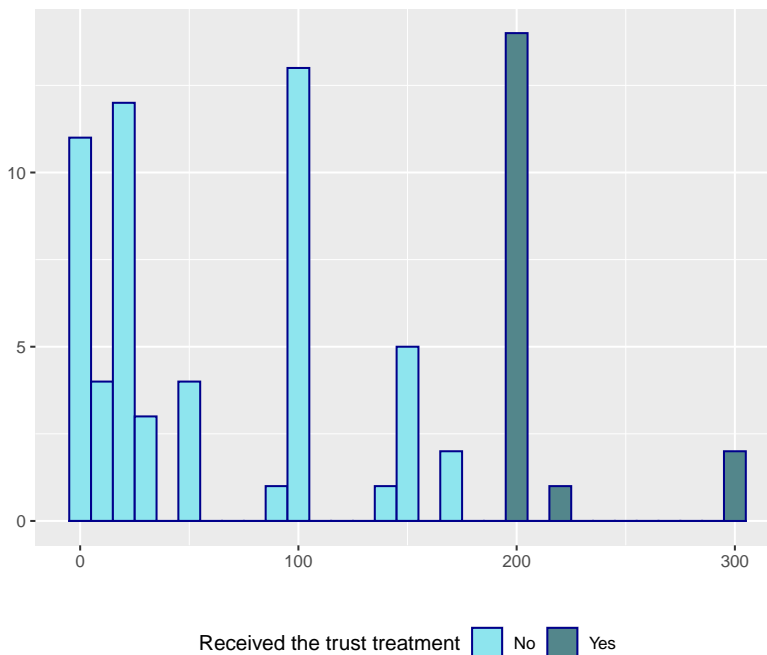


Figure 7 – Histogram of share of points sent back in the trust game

### 3.c.2 Variables of participation to the questionnaire

As said above, 5 variables measure participation to the questionnaire : the number of close-ended and open-ended questions answered, the number of words used in the open-ended questions (total word count), the average number of words used per open-ended question, and the total duration of the experiment. Table 18 provides some summary statistics of these variables.

Statistic	N	Mean	St. Dev.	Min	Median	Max
Close-ended questions	73	2.863	0.608	0	3	3
Open-ended questions	73	2.630	2.105	0	3	5
Total word count	50	65.200	55.507	2	45.500	237
Average word count	50	16.004	10.794	1.667	13.375	47.400
Duration (in seconds)	73	593.068	600.528	205	397	4,500

Table 18 – Summary statistics of the participation variables of interest

Almost all subjects answered to all three close-ended questions, except for 3 subjects who answer to none and one who answers to only 2. Therefore, this variable does not exhibit sufficient variation to help discriminate subjects based on their participation to the questionnaire, and it is not being considered in the subsequent regression analysis. However, the finding that the number of close-ended questions answered is not responsive to the level of interpersonal trust in this experiment is actually in line with the cross-sectional geographic evidence from section 2.d.

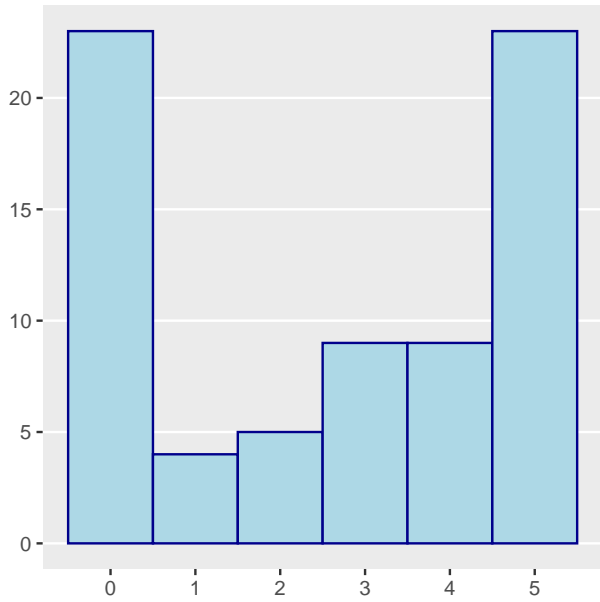
There is more heterogeneity in the number of open-ended questions answered. Roughly one third of subjects (23 among 73) answered to no open-ended questions, one-third (23) responded to every open-ended question, and one-third (27) answered to some but not all of the open-ended questions (see figure 8a).

The distribution of the experiment duration variable has a right tail (see figure 8b) : the experiment took around 7 minutes or less (the median duration is 397 seconds, i.e. 6 minutes and 37 seconds) for half of the subjects, but it took significantly longer for a fraction of subjects. When analyzing the duration variable, 3 outlying values are removed from the analysis, since it can be assumed that subjects whose experiment lasted more than 30 minutes (1800 seconds) are clearly subjects who did not complete the experiment in one stroke and therefore cannot be compared to the other subjects.

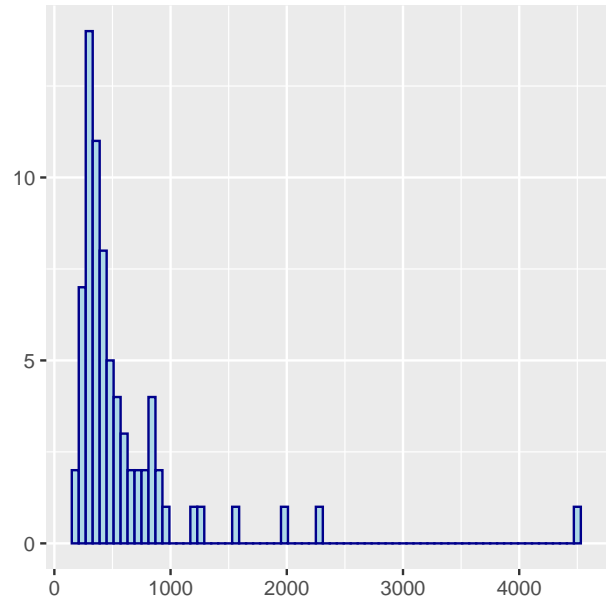
In terms of the number of words used in open-ended questions, most subjects answer to each question with one or two sentences, i.e. 10 to 20 words (see figure 8c and 8d), sometimes more, but no striking outlying value stands out<sup>11</sup>. Manual check shows that no gibberish language was used, which confirms the good quality of the subset of subjects who passed the attention check.

These four questionnaire variables of interest are all positively correlated to one another (see figure 9). This result is in line with the idea that all these variables measure the subject’s latent willingness to participate to the questionnaire.

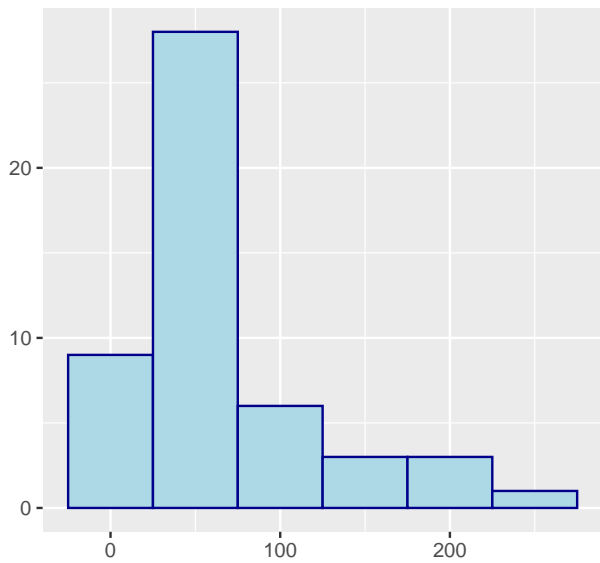
<sup>11</sup>The longest answer to a single open-ended questions contains 77 words.



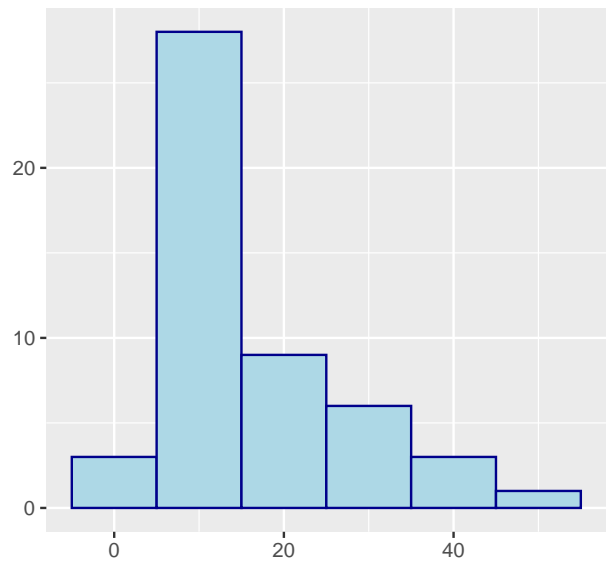
(a) Number of open-ended questions answered



(b) Duration (in seconds, binwidth = 60 seconds)



(c) Total word count (for subjects answering to at least one question)



(d) Average number of words per open-ended question (for subjects answering to at least one question)

Figure 8 – Histograms of questionnaire variables of interest

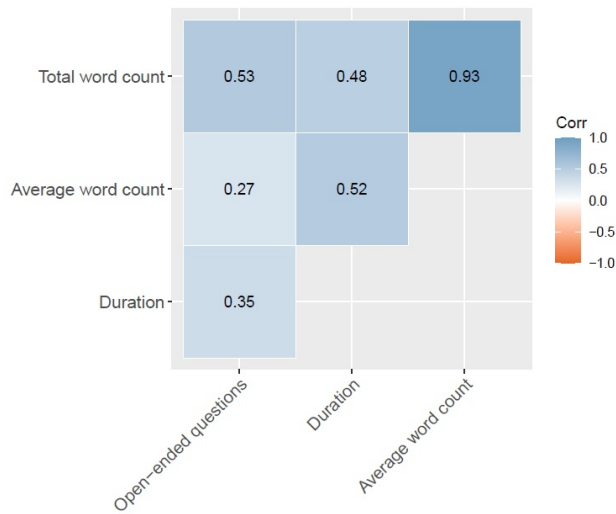


Figure 9 – Correlation matrix of the questionnaire variables of interest

### 3.d Results

To assess jointly the two hypotheses formulated in subsection 3.b.2, the following regression specification is considered:

$$Y_i = \alpha + \beta T_i + \gamma M_i + \epsilon_i$$

where  $Y_i$  is an outcome variable of participation to the questionnaire,  $T_i$  is the indicator variable equal to 1 when the amount sent in the trust game is higher or equal to 7 (“trusting” individual), and  $M_i$  is the indicator variable equal to 1 when the amount sent back in the trust game is more than twice the amount sent (“trust treatment”).

When considering the total or average number of words as the outcome variable, this regression is run on the subsample of 50 subjects who answered to at least one open-ended question. When considering the total duration of the experiment, 3 outlying values are removed as mentioned above.

The results are displayed in table 19. Standard errors are bootstrapped to account for the small sample size.

	<i>Dependent variable:</i>			
	Open-ended questions	Total word count	Average word count	Duration
	(1)	(2)	(3)	(4)
Trusting individual	1.09* (0.614)	47.05*** (16.18)	6.07* (3.49)	118.87* (72.08)
Trust treatment	0.97* (0.581)	30.00* (17.71)	5.18 (3.74)	62.21 (94.73)
Observations	73	50	50	70

Note: bootstrapped standard errors (n=1000)

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 19 – Regressions of participation variables of interest on trust game variables

Being a “trusting” individual, on the second hump of the distribution of the amounts sent in the trust game, leads to more participation in the questionnaire along the four dimensions considered. Trusting individuals answer to one more open-ended question, and use 6 more words per open-ended question, which translates into 47 more words used in total (almost one standard deviation). They spend two more minutes taking part in the experiment.

The trust treatment only leads to more participation in terms of open-ended questions answered (one more open-ended questions answered) and total number of words used in the open-ended questions.

Despite its limited scale, this pilot experiment confirms the baseline hypothesis that trust drives voluntary participation at the intensive margin to an online questionnaire on topical issues. More trusting people, as elicited by the amount they send in the trust game played with a stranger, tend to participate more to the questionnaire akin to a public consultation. Trust manipulation, through the random amounts which are sent back in the trust game, plays a role as well in driving more participation. Being sent back a higher amount is a random treatment that makes the subject participate more to the questionnaire, for a given initial level of trust, by boosting the subject’s trust in others.

These results hold when considering alternative thresholds for defining trusting individuals and the trust treatment, or when considering directly the continuous variables “Amount sent in the game” and “Share sent back” : the point estimates are in the same direction, but they may not achieve statistical significance (see appendix table 40 for regression results with continuous variables instead of binary variables).

### **3.e Remarks and limitations**

This experiment faces some limitations. In particular, the causal effect of trust manipulation via the random amount sent back in the trust game could also be driven by other channels. A higher amount which is sent back is likely to drive the subject’s level of trust in others, as I explained, but can also, for instance, simply act on the subject’s overall mood (winning points makes people happy), which can also lead to more participation to the questionnaire. The existence of these other channels is possible, but is hard to control for in a small-scale experiment.

Building on this pilot experiment, possible improvements include the extension of this experiment to more participants to be better able to assess the statistical significance of our results, as well as robustness checks using alternative measures of trust such as the ones described in [Glaeser et al. \(2000\)](#). Another practical improvement would be to set up a time limit to avoid having to discard subjects whose experiment duration was too long.

## 4 Conclusion

This work helps shed light on the role of interpersonal trust in driving behaviours with regards to participation to public consultations, especially online, by providing a variety of evidence, both geographic cross-sectional evidence from the study of *Grand Débat National* as well as experimental evidence.

My main finding is that interpersonal trust makes people participate more to participatory democracy, at the extensive and intensive margins. This finding is backed by cross-sectional geographic evidence from *Grand Débat National*, as well as by experimental evidence. This effect is stronger for open-ended questions than for close-ended questions : more trusting individuals will answer more open-ended questions than less trusting individuals, but no difference exists when it comes to close-ended questions.

These results implies that participatory democracy tools such as public consultations will tend to be taken over, metaphorically speaking, by trusting individuals, while less trusting individuals will remain relatively absent from them. This effect will be aggravated when these consultations include open-ended questions. Therefore, far from being a solution to tackle the estrangement from politics by low-trust individuals, public consultations run the risk of giving even more voice to high-trust individuals instead. In the context of the *Gilets Jaunes* movement in France, where the *Gilets Jaunes* protesters had strikingly low levels of interpersonal trust, it would appear that they remained absent from the *Grand Débat National* though it was originally designed to give them a voice. To account for the success of the *Grand Débat National*, it can be assumed that what happened is that other, more trusting, individuals, took advantage of this participatory democratic tool to express themselves to influence policy-making.

The practical implications of this differential level of participation between low-trust and high-trust individuals should be minimal if all individuals expressed themselves similarly during public consultations. Indeed, if the government takes its decisions based on the results of these public consultations and if low-trust individuals are simply less present, but expressing the same demands as high-trust individuals, then it should not matter too much for individuals with lower levels of interpersonal trust, since the policies they support will be implemented.

However, in the case of the *Grand Débat National*, I find some striking discrepancies in the content of contributions, along various dimensions, depending on the level of interpersonal trust. Between low-trust and high-trust life zones, the tone is not the same, the complexity of the language is different, as well as the content of the contributions themselves, as captured by the profiles of respondents and the words used. This hints at the existence of a gap in terms of what low-trust and high-trust individuals aimed to express in this public consultation. In that case, the fact that public consultations tend to be dominated by the contributions of high-trust individuals leads to the overrepresentation of their demands.

In a dynamic setting where the government adjusts policies in reaction to public consultations, this will tend to accentuate the disparities between high-trust and low-trust individuals in their control over policy-making, possibly aggravating the political polarization of societies along the high-low trust divide : on the one hand, high-trust individuals will have an increased control over political decisions, while low-trust individuals will be further relegated away from political decisions. The process could even be self-sustaining, if trust is itself an increasing function of the power that each citizen has (or perceives to have) over political decisions. The following process could be going on : individuals with lower levels of interpersonal trust abstain from a given public consultation, which is thus flooded by individuals with higher levels of interpersonal trust. As a result, the policies implemented based on this public consultation are favoring people who participated over those who abstained. THE feeling to have little influence over political decisions then drives down the trust of people who abstained, who already had lower levels of trust to begin with. This dynamic effect of public consultations, and participatory democracy in general, on interpersonal trust and political behaviours, calls for a possible new line of research which is out of the scope of this work.



This work also puts forth a number of methodological contributions :

- Using spatial micro-simulation to study the geographic distribution of certain variables, such as interpersonal trust, and use this distribution to study its cross-sectional relationship with some variables of interest. In the absence of relevant microdata and whenever a phenomenon has a geographical dimension, this is an attractive method to study correlations between variables, despite the impossibility to conclude on clear causal inference.
- Profile building using natural language processing machine learning tools, namely Latent Dirichlet Allocation (LDA), applied on large-scale survey questionnaires, including close-ended as well as open-ended questions. Such methods could be implemented in other contexts, whenever multidimensional data has to be summarized into a limited set of dimensions in an inductive manner.
- A tractable experimental framework to study the role of interpersonal trust on participation to public consultations.

These results are encouraging and offer multiple ways ahead to build on and improve them going forward. First of all, the cross-sectional and experimental results presented are for the most part correlational and do not point out the causal relationship between interpersonal trust and participation to public consultations. If individual microdata on participation to *Grand Débat National* or other public consultations is to be available one day, then this would call for multivariate regression analysis that identifies the effect of interpersonal trust, holding other possible confounders such as age, education, political affiliation and gender, constant. Another way ahead would be to measure trust using alternative methods : both experimentally, inspired by [Glaeser et al. \(2000\)](#) where several alternative methods are put forward, and also in the cross-sectional analysis.

In a more ambitious way, the predictive role of interpersonal trust in explaining democratic participation should also be studied in other instances : from [Algan et al. \(2018, 2019b\)](#), it appears that *Gilets Jaunes* activity is negatively related with interpersonal trust, while the Macron vote and participation to public consultations are positively related with interpersonal trust. What about other forms of participatory democracy, such as participatory budgeting assemblies, citizens' juries or citizens' assemblies? What about online petitioning, demonstrating, voter turnout, or localized online Facebook activity ([Boyer et al., 2020](#))? Studying how these forms of political expression interact with interpersonal trust would enable to compare which democratic practice is most complementary with interpersonal trust, which is not and build a spectrum of the forms of democratic expression and participation for different levels of interpersonal trust. If interpersonal trust proves to be a key predictor for these behaviours, then this work would surely help shed light on the workings of political behaviours and expression.

## 5 Appendices

### 5.a List of questions from Grand Débat National

#### 5.a.1 Short questionnaires

Question Text	Possible answers
Selon vous, faut-il introduire une dose de proportionnelle pour certaines élections, lesquelles ?	Régionales, Départementales, Législatives, Il ne faut pas introduire de proportionnelle <i>Answers can be combined</i>
Pensez-vous qu'il serait souhaitable de réduire le nombre de parlementaires (députés + sénateurs = 925) ?	Oui, Non
Faut-il rendre le vote obligatoire ?	Oui, Non
Faut-il avoir davantage recours au référendum au niveau national ?	Oui, Non
Faut-il avoir davantage recours au référendum au niveau local ?	Oui, Non
Faut-il tirer au sort des citoyens non élus pour les associer à la décision publique ?	Oui, Non, Je ne sais pas
Diriez-vous que l'application de la laïcité en France est aujourd'hui:	A améliorer, à modifier profondément, Satisfaisante

Table 20 – Democracy and citizenship (short questionnaire)

<b>Question Text</b>	<b>Possible answers</b>
Pensez-vous que vos actions en faveur de l'environnement peuvent vous permettre de faire des économies ?	Oui, Non
Diriez-vous que vous connaissez les aides et dispositifs qui sont aujourd'hui proposés par l'Etat, les collectivités, les entreprises et les associations pour l'isolation et le chauffage des logements, et pour les déplacements ?	Oui, Non
Pensez-vous que les taxes sur le diesel et sur l'essence peuvent permettre de modifier les comportements des utilisateurs ?	Oui, Non
À quoi les recettes liées aux taxes sur le diesel et l'essence doivent-elles avant tout servir ?	À baisser d'autres impôts comme par exemple l'impôt sur le revenu, À financer des aides pour accompagner les Français dans la transition écologique, À financer des investissements en faveur du climat
Selon vous, la transition écologique doit être avant tout financée :	Par la fiscalité écologique, Par le budget général de l'État, Les deux, Je ne sais pas
Et qui doit être en priorité concerné par le financement de la transition écologique ?	Les administrations, Les entreprises, Les particuliers, Tout le monde <i>Answers can be combined</i>
Que faudrait-il faire pour protéger la biodiversité et le climat tout en maintenant des activités agricoles et industrielles compétitives par rapport à leurs concurrents étrangers, notamment européens ?	Cofinancer un plan d'investissement pour changer les modes de production, Modifier les accords commerciaux, Taxer les produits importés qui dégradent l'environnement

Table 21 – Ecological transition (short questionnaire)

<b>Question Text</b>	<b>Possible answers</b>
Afin de réduire le déficit public de la France qui dépense plus qu'elle ne gagne, pensez-vous qu'il faut avant tout:	Augmenter les impôts, Réduire la dépense publique, Faire les deux en même temps, Je ne sais pas
Afin de baisser les impôts et réduire la dette, quelles dépenses publiques faut-il réduire en priorité ?	Les dépenses des collectivités territoriales, Les dépenses de l'Etat, Les dépenses sociales, Je ne sais pas
Parmi les dépenses de l'Etat et des collectivités territoriales, dans quels domaines faut-il faire avant tout des économies ?	<i>Open-ended question</i>
Seriez-vous prêt à payer un impôt pour encourager des comportements bénéfiques à la collectivité comme la fiscalité écologique ou la fiscalité sur le tabac ou l'alcool ?	Oui, Non

Table 22 – Fiscality and public spending (short questionnaire)

Question Text	Possible answers
Savez-vous quels sont les différents échelons administratifs (Etat, collectivités territoriales comme la région, la commune, opérateurs comme par exemple Pole Emploi ou la CAF) qui gèrent les différents services publics dans votre territoire ?	Oui, Non
Pensez-vous qu'il y a trop d'échelons administratifs en France ?	Oui, Non
Quels sont les niveaux de collectivités territoriales auxquels vous êtes le plus attaché ?	Commune, Intercommunalité, Département, Région <i>Answers can be combined</i>
Lorsqu'un déplacement est nécessaire pour effectuer une démarche administrative, quelle distance pouvez-vous parcourir sans difficulté ?	Jusqu'à 5 kms, Jusqu'à 10 kms, Jusqu'à 15 kms, Jusqu'à 20 kms, Plus de 20 kms
Pour accéder à certains services publics, vous avez avant tout des besoins...	Numériques, Physiques pour pouvoir vous rendre sur place, Téléphoniques <i>Answers can be combined</i>
Si vous rencontrez des difficultés pour effectuer vos démarches administratives sur Internet, de quel accompagnement souhaiteriez-vous bénéficier ?	Une formation numérique, Une aide téléphonique, Une prise en charge par un agent <i>Answers can be combined</i>
Seriez-vous d'accord pour qu'un agent public effectue certaines démarches à votre place ?	Oui, Non
Que pensez-vous du regroupement dans un même lieu de plusieurs services publics (Maisons de services au public) ?	Bonne chose, Mauvaise chose
Que pensez-vous des services publics itinérants (bus de services publics) ?	Bonne chose, Mauvaise chose
Que pensez-vous du service public sur prise de rendez-vous ?	Bonne chose, Mauvaise chose
Que pensez-vous des agents publics polyvalents susceptibles de vous accompagner dans l'accomplissement de plusieurs démarches quelle que soit l'administration concernée ?	Bonne chose, Mauvaise chose
Avez-vous déjà renoncé à des droits / des allocations en raison de démarches administratives trop complexes ?	Oui, Non

Table 23 – Organisation of the State and public services (short questionnaire)

## 5.a.2 Long questionnaires

Question Text	Possible answers	Keywords for profile construction using LDA
En qui faites-vous le plus confiance pour vous faire représenter dans la société et pourquoi ?	<i>Open-ended question</i>	Maire / élu municipal ; Député ("député", "assemblée Nationale") ; Sénateur ; Président de la République ("président", "macron") ; Association ; Syndicat ; Personne ("personne", "moi-même") ; Elus locaux ; Elus en général
En dehors des élus politiques, faut-il donner un rôle plus important aux associations et aux organisations syndicales et professionnelles ?	Oui, Non	<i>Same as suggested answers</i>
Si oui, à quel type d'associations ou d'organisations ? Et avec quel rôle ?	<i>Open-ended question</i>	Syndicat ; Société civile ("civil", "citoyen") ; Association ; Association de consommateur ("consommateur") ; Corps intermédiaire
Que faudrait-il faire pour renouer le lien entre les citoyens et les élus qui les représentent ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Le non-cumul des mandats instauré en 2017 pour les parlementaires (députés et sénateurs) est :	Une bonne chose, Une mauvaise chose, Je ne sais pas	<i>Same as suggested answers</i>
Pourquoi ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Que faudrait-il faire pour mieux représenter les différentes sensibilités politiques ?	<i>Open-ended question</i>	Proportionnelle Référendum ("référendum", "RIC") Vote blanc ("blanc")
Pensez-vous qu'il serait souhaitable de réduire le nombre d'élus (hors députés et sénateurs) ?	Oui, Non	<i>Same as suggested answers</i>
Si oui, lesquels ?	<i>Open-ended question</i>	Sénateur ; Député ; Elus régionaux ("région") ; Elus municipaux ("commune", "municipal") ; Intercommunalité ; Elus départementaux ("département", "canton")
Que pensez-vous de la participation des citoyens aux élections et comment les inciter à y participer davantage ?	<i>Open-ended question</i>	Vote obligatoire ("vote obligatoire", "amende") Vote blanc ("blanc") Proportionnelle Education Vote en ligne ("ligne", "internet", "électronique") Référendum ("référendum", "RIC")
Faut-il prendre en compte le vote blanc ?	Oui, Non	<i>Same as suggested answers</i>
Si oui, de quelle manière ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Que faudrait-il faire aujourd'hui pour mieux associer les citoyens aux grandes orientations et à la décision publique ?	<i>Open-ended question</i>	Référendum ("référendum", "RIC", "rIP")
Faut-il faciliter le déclenchement du référendum d'initiative partagée (le RIP est organisé à l'initiative de membres du Parlement soutenu par une partie du corps électoral) qui est applicable depuis 2015 ?	Oui, Non, Je ne sais pas	<i>Same as suggested answers</i>
Si oui, comment ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Que faudrait-il faire pour consulter plus directement les citoyens sur l'utilisation de l'argent public, par l'Etat et les collectivités ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Quel rôle nos assemblées, dont le Sénat et le Conseil économique, social et environnemental, doivent-elles jouer pour représenter nos territoires et la société civile ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Faut-il les transformer ?	Oui, Non	<i>Same as suggested answers</i>
Si oui, comment ?	<i>Open-ended question</i>	Supprimer le Sénat Supprimer le CESE Diminuer le nombre de sénateurs Changer le mode d'élection des sénateurs

Table 24 – Democracy and citizenship (long questionnaire) (1/2)

Question Text	Possible answers	Keywords for profile construction using LDA
Que proposez-vous pour renforcer les principes de la laïcité dans le rapport entre l'Etat et les religions de notre pays ?	<i>Open-ended question</i>	Loi de 1905 ("1905") Respect ("respect", "tolérance", "vivre-ensemble") Education ("éducation", "école") Autorité ("interdire", "justice", "amende")
Comment garantir le respect par tous de la compréhension réciproque et des valeurs intangibles de la République ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Que faudrait-il faire aujourd'hui pour renforcer l'engagement citoyen dans la société ?	<i>Open-ended question</i>	Education ("éducation", "école") Service civique Service militaire Association
Quels sont les comportements civiques qu'il faut promouvoir dans notre vie quotidienne ou collective ?	<i>Open-ended question</i>	Respect ("respect", "vivre-ensemble", "politesse", "règles", "civilité") Solidarité ("solidarité", "entraide", "association", "aide", "égalité", "engagement", "bénévole") Environnement ("environnement", "pollution", "déchet")
Que faudrait-il faire pour favoriser le développement de ces comportements civiques et par quels engagements concrets chacun peut-il y participer ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Que faudrait-il faire pour valoriser l'engagement citoyen dans les parcours de vie, dans les relations avec l'administration et les pouvoirs publics ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Quelles sont les incivilités les plus pénibles dans la vie quotidienne et que faudrait-il faire pour lutter contre ces incivilités ?	<i>Open-ended question</i>	Respect ("respect", "politesse", "courtoisie", "civisme") Pollution ("déchet", "pollution", "décharge", "mégot", "environnement") Dégradation ("dégradation", "casse", "sécurité", "agression") Racisme
Que peuvent et doivent faire les pouvoirs publics pour répondre aux incivilités ?	<i>Open-ended question</i>	Sanction ("amende", "sanction", "justice", "interdire", "travaux d'intérêt général") Education ("éducation", "école", "pédagogie")
Quel pourrait être le rôle de chacun pour faire reculer les incivilités dans la société ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Quelles sont les discriminations les plus répandues dont vous êtes témoin ou victime ?	<i>Open-ended question</i>	Racisme Sexisme ("sexisme", "femme") Homophobie Racisme anti-blanc ("blanc") Discrimination contre les personnes handicapées ("handicap") Antisémitisme Discrimination contre les musulmans Discrimination contre les chrétiens / catholiques Discrimination contre les personnes âgées ("âge", "retraite")
Que faudrait-il faire pour lutter contre ces discriminations et construire une société plus solidaire et plus tolérante ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Pensez-vous qu'il faille instaurer des contreparties aux différentes allocations de solidarité ?	Oui, Non	<i>Same as suggested answers</i>
Si oui, lesquelles ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Que pensez-vous de la situation de l'immigration en France aujourd'hui et de la politique migratoire ? Quelles sont, selon vous, les critères à mettre en place pour définir la politique migratoire ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
En matière d'immigration, une fois nos obligations d'asile remplies, souhaitez-vous que nous puissions nous fixer des objectifs annuels définis par le Parlement ?	<i>Open-ended question</i>	Oui Non Quota
Que proposez-vous afin de répondre à ce défi qui va durer ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>
Quelles sont, selon vous, les modalités d'intégration les plus efficaces et les plus justes à mettre en place aujourd'hui dans la société ?	<i>Open-ended question</i>	Education ("éducation", "école") Travail ("travail", "formation", "stage") Langue ("langue", "parler", "linguistique") Valeur ("valeur", "respect") Logement
Y a-t-il d'autres points sur la démocratie et la citoyenneté sur lesquels vous souhaiteriez vous exprimer ?	<i>Open-ended question</i>	<i>Excluded from the analysis</i>

Table 25 – Democracy and citizenship (long questionnaire) (2/2)

Question Text	Possible answers
Quel est aujourd'hui pour vous le problème concret le plus important dans le domaine de l'environnement ?	Pollution de l'air Erosion du littoral Dérèglements climatiques (crues, sécheresses) Biodiversité et disparition de certaines espèces + Open text box
Que faudrait-il faire selon vous pour apporter des réponses à ce problème ?	<i>Open-ended question</i>
Diriez-vous que votre vie quotidienne est aujourd'hui touchée par le changement climatique ?	Oui, Non
Si oui, de quelle manière votre vie quotidienne est-elle touchée par le changement climatique ?	<i>Open-ended question</i>
Si oui, que faites-vous aujourd'hui pour protéger l'environnement et/ou que pourriez-vous faire ?	<i>Open-ended question</i>
Qu'est-ce qui pourrait vous inciter à changer vos comportements comme par exemple mieux entretenir et régler votre chauffage, modifier votre manière de conduire ou renoncer à prendre votre véhicule pour de très petites distances ?	<i>Open-ended question</i>
Quelles seraient pour vous les solutions les plus simples et les plus supportables sur un plan financier pour vous inciter à changer vos comportements ?	<i>Open-ended question</i>
Par rapport à votre mode de chauffage actuel, pensez-vous qu'il existe des solutions alternatives plus écologiques ?	Oui, Non
Si oui, que faudrait-il faire pour vous convaincre ou vous aider à changer de mode de chauffage ?	<i>Open-ended question</i>
Avez-vous pour vos déplacements quotidiens la possibilité de recourir à des solutions de mobilité alternatives à la voiture individuelle comme les transports en commun, le covoiturage, l'auto-partage, le transport à la demande, le vélo, etc. ?	Oui, Non, Je n'utilise pas la voiture pour les déplacements quotidiens
Si oui, que faudrait-il faire pour vous convaincre ou vous aider à utiliser ces solutions alternatives ?	<i>Open-ended question</i>
Si non, quelles sont les solutions de mobilité alternatives que vous souhaiteriez pouvoir utiliser ?	Transports en commun Covoiturage Vélo Transport à la demande Auto-partage + Open text box
Et qui doit selon vous se charger de vous proposer ce type de solutions alternatives ?	<i>Open-ended question</i>
Que pourrait faire la France pour faire partager ses choix en matière d'environnement au niveau européen et international ?	<i>Open-ended question</i>
Y a-t-il d'autres points sur la transition écologique sur lesquels vous souhaiteriez vous exprimer ?	<i>Open-ended question</i>

Table 26 – Ecological transition (long questionnaire)

Question Text	Possible answers
Quelles sont toutes les choses qui pourraient être faites pour améliorer l'information des citoyens sur l'utilisation des impôts ?	<i>Open-ended question</i>
Que faudrait-il faire pour rendre la fiscalité plus juste et plus efficace ?	<i>Open-ended question</i>
Quels sont selon vous les impôts qu'il faut baisser en priorité ?	<i>Open-ended question</i>
Afin de financer les dépenses sociales, faut-il selon vous...	Augmenter le temps de travail, Reculer l'âge de la retraite, Revoir les conditions d'attribution des aides sociales, Augmenter les impôts + Open text box
S'il faut selon vous revoir les conditions d'attribution de certaines aides sociales, lesquelles doivent être concernées ?	<i>Open-ended question</i>
Quels sont les domaines prioritaires où notre protection sociale doit être renforcée ?	<i>Open-ended question</i>
Pour quelle(s) politique(s) publique(s) ou pour quels domaines d'action publique, seriez-vous prêts à payer plus d'impôts ?	<i>Open-ended question</i>
Y a-t-il d'autres points sur les impôts et les dépenses sur lesquels vous souhaiteriez vous exprimer ?	<i>Open-ended question</i>

Table 27 – Fiscality and public spending (long questionnaire)



Question Text	Possible answers
Que pensez-vous de l'organisation de l'Etat et des administrations en France ? De quelle manière cette organisation devrait-elle évoluer ?	<i>Open-ended question</i>
Selon vous, l'Etat doit-il aujourd'hui transférer de nouvelles missions aux collectivités territoriales ?	Oui, Non
Si oui, lesquelles ?	<i>Open-ended question</i>
Estimez-vous avoir accès aux services publics dont vous avez besoin ?	Oui, Non
Si non, quels types de services publics vous manquent dans votre territoire et qu'il est nécessaire de renforcer ?	<i>Open-ended question</i>
Quels nouveaux services ou quelles démarches souhaitez-vous voir développées sur Internet en priorité ?	<i>Open-ended question</i>
Avez-vous déjà utilisé certaines de ces nouvelles formes de services publics ?	Oui, Non
Si oui, en avez-vous été satisfait ?	Oui, Non
Quelles améliorations préconiserez-vous ?	<i>Open-ended question</i>
Quand vous pensez à l'évolution des services publics au cours des dernières années, quels sont ceux qui ont évolué de manière positive ?	<i>Open-ended question</i>
Quels sont les services publics qui doivent le plus évoluer selon vous ?	<i>Open-ended question</i>
Connaissez-vous le "droit à l'erreur", c'est-à-dire le droit d'affirmer votre bonne foi lorsque vous faites un erreur dans vos déclarations ?	Oui, Non
Si oui, avez-vous déjà utilisé ce droit à l'erreur ?	Oui, Non
Si oui, à quelle occasion en avez-vous fait usage ?	<i>Open-ended question</i>
Pouvez-vous identifier des règles que l'administration vous a déjà demandé d'appliquer et que vous avez jugées inutiles ou trop complexes ?	<i>Open-ended question</i>
Faut-il donner plus d'autonomie aux fonctionnaires de terrain ?	Oui, Non
Si oui, comment ?	<i>Open-ended question</i>
Faut-il revoir le fonctionnement et la formation de l'administration ?	Oui, Non
Si oui, comment ?	<i>Open-ended question</i>
Comment l'Etat et les collectivités territoriales peuvent-ils s'améliorer pour mieux répondre aux défis de nos territoires les plus en difficulté ?	<i>Open-ended question</i>
Si vous avez été amené à chercher une formation, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à scolariser votre enfant, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à chercher un emploi, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à préparer votre retraite, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à demander un remboursement de soins de santé, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à faire une demande d'aide pour une situation de handicap, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à créer une entreprise, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à recruter du personnel, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à former du personnel, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à rémunérer du personnel, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à mettre fin à votre activité, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Si vous avez été amené à recruter une personne portant un handicap, pouvez-vous indiquer les éléments de satisfaction et/ou les difficultés rencontrés en précisant, pour chaque point, l'administration concernée :	<i>Open-ended question</i>
Y a-t-il d'autres points sur l'organisation de l'Etat et des services publics sur lesquels vous souhaiteriez vous exprimer ?	<i>Open-ended question</i>

Table 28 – Organisation of the State and public services (long questionnaire)

## 5.b Spatial micro-simulation : method, construction, and validity tests

### 5.b.1 Method : synthetic micro-simulation using the iterative proportional fitting

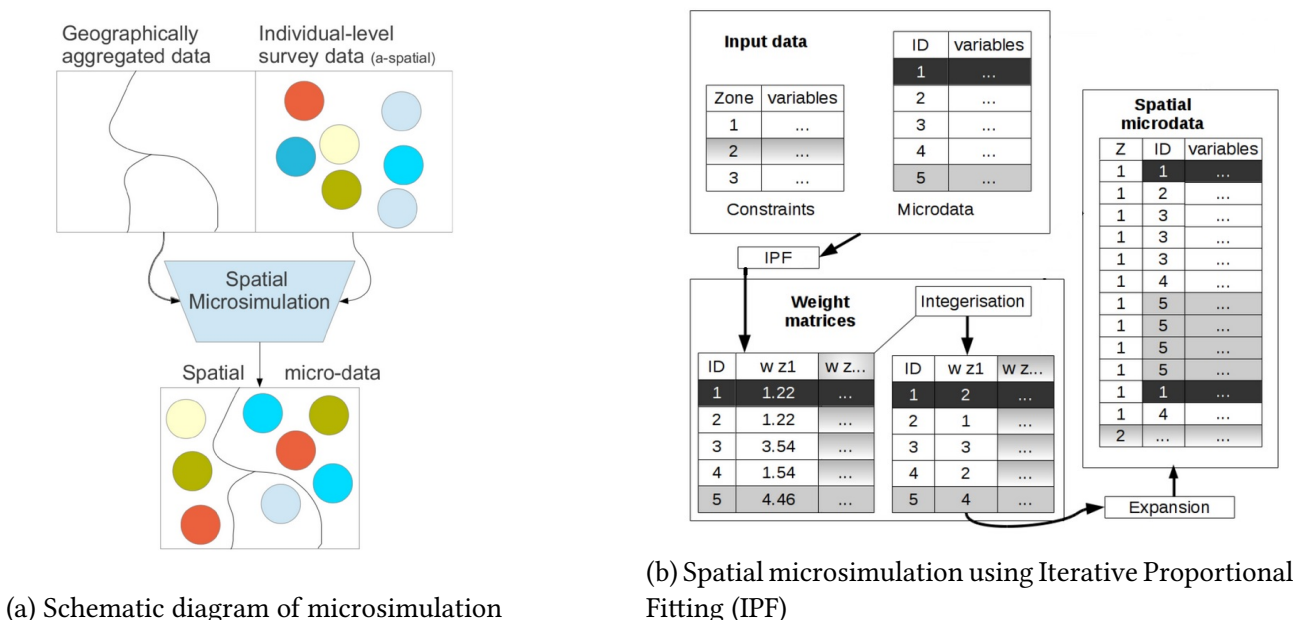


Figure 10 – Diagrams from Lovelace and Dumont (2016)

In short, spatial microsimulation is a form of population synthetis, whereby individuals from survey data are allocated to geographical units using geographically aggregated data (see figure 10a). To be performed, this method thus relies on two sets of data :

- Geographic aggregated data, containing a number of constraint variables. These variables are the counts of certain types of individuals. For example, it says that in a given life zone, there are 9 individuals : 4 women and 5 men, 3 individuals above 50 and 6 below 50.
- Survey individual microdata, which contains variables of two kinds : linking variables matching the constraint variables of the geographic aggregated data (in this example, gender and age), and the target variable whose geographic distribution we want to generate (interpersonal trust in our case).

I follow the three-step method described in Lovelace and Dumont (2016) and summarized in figure 10b. The first step is to use Iterative Proportional Fitting (IPF) : for each life zone, it consists in assigning weights to observations from the individual microdata to get the weight matrix that matches best the characteristics of the population of this life zone, as given by the constraint variables. In this example, it would assign weights to observations in the microdataset to have a weighted synthetic population with as close as 4 women, 5 men, 3 individuals below 50 and 6 individuals above 60.

The next step is integerisation : the weights computed through this procedure are real numbers, through integerisation, they are rounded into integer numbers. Integer weights can be interpreted in a straightforward manner as the number of such individuals assigned into the synthetic population of the life zone. Integerisation can be implemented using a variety of techniques, I rely on the “Truncate, Replicate, Sample” method put forward in Lovelace and Ballas (2013).

The final step is expansion : the weight matrix produced by integerisation is converted into the final spatial microdata output in “long” format. Each line is an individual from the input microdata assigned to a given

life zone. Note that each individual can possibly be assigned multiple times to a given life zone, or to different life zones. Since the level of interpersonal trust of each individual in the synthetic population is known (from the survey microdata), averaging these values enable to compute the (average) level of local interpersonal trust of a life zone. The resulting variable will be referred to as the local interpersonal trust of a given life zone.

### **5.b.2 Construction of the local interpersonal trust variable with synthetic micro-simulation**

To build the geographic aggregated data used for the spatial microsimulation procedure described above, I rely on 2016 census data from the French National Statistical Institute (INSEE) giving the cross-tabulation of gender, education level (4 modalities) and age (5 modalities) at the municipality level (40 modalities overall) ; and on the French Interior Ministry data on the election results for the first round of the 2017 presidential at the municipality level (7 modalities). Appendix section 5.b.6 summarizes details on the modalities. Both datasets are merged and variables are aggregated at the life zone level.

The survey individual microdata used is the 2017 wave of the French Electoral Survey (CEVIPOF), which collected data from a representative sample of 1830 French respondents above 18. Crucially, as highlighted in [Lovell and Dumont \(2016\)](#), the individual-level data used is representative of the study area, metropolitan France in this case, which is a key assumption for spatial microsimulation to be valid. The linking variables are the age, gender, level of education, and vote in the first round of the 2017 presidential election and their modalities are harmonized with the ones from geographic aggregated data. This dataset includes our target variable, interpersonal trust, measured with the standard question “On a 0 to 10 scale, would you say that you can trust most people or that one is never cautious enough when treating with others ?”.

Preliminary cleaning includes the removal of 60 respondents of ages 18 and 19 from the individual-level data, as well as 3 respondents who answer “Don’t know” to the interpersonal trust question (1767 remaining observations). Similarly, any individual below 20 from geographical-level data are removed to get consistent datasets. The cost in terms of accuracy for the synthetic variable is likely to be minimal. Appendix table 32 provides summary statistics of the survey individual microdata for the linking variables. Table 33 provides summary statistics for the interpersonal trust variable in this sample.

The constraint variables for the calibration of the spatial microsimulation IPF algorithm (age, gender, education, and vote in 2017 Presidential election) are chosen after [Algan et al. \(2018\)](#), which highlights their predictive role to explain the level interpersonal trust, and uses them to build a synthetic local trust variable.

It is worth noting that this approach rests on the assumption that individual socio-demographic and political characteristics are the main determinants of trust. In fact, trust depends on a variety of other characteristics as evidenced in [Alesina and La Ferrara \(2002\)](#) : individual and group past experience, as well as local community characteristics. However, these variables are not available either in the geographic or individual datasets, hence they cannot be included in the construction of this synthetic microsimulated interpersonal trust variable.

Table 29 highlights the predictive role of the chosen constraint variables to explain interpersonal trust in the individual microdataset. The low value for the  $R^2$ , consistent with [Alesina and La Ferrara \(2002\)](#), underlines the fact that many determinants of trust are not captured.

	<i>Dependent variable:</i>
	Interpersonal trust (0-10 scale)
Female	-0.203* (0.116)
Age	0.017 (0.018)
Age <sup>2</sup>	-0.00001 (0.0002)
Secondary vocational diploma (CAP, BEP)	0.067 (0.173)
High school diploma	0.524*** (0.194)
Higher education diploma	1.066*** (0.180)
Abstention	0.306 (0.208)
Fillon voter	0.748*** (0.228)
Hamon voter	1.972*** (0.289)
Macron voter	1.193*** (0.201)
Mélenchon voter	1.220*** (0.200)
Other voter	0.506** (0.209)
Constant	2.189*** (0.477)
Observations	1,767
R <sup>2</sup>	0.101

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Reference categories : Male, No diploma, Le Pen voter

Table 29 – Regression of interpersonal trust on the constraints variables in the individual microdataset

### 5.b.3 Internal validation

Lovelace and Dumont (2016) provide two tests for internal validation of the spatial microsimulation approach.

The first one is to check for a possible empty cell problem, which is not the case here. This problem appears when the linking variables are defined at a too fine level of granularity, which entails that there is no individual possessing certain characteristics in the individual microdataset (for instance, the sample does not include any male of age 20-22 with a higher education diploma). This problem is alleviated through the careful choice of the modalities of the linking variables to avoid having too small categories (in that case, broadening the age variable modalities to 20-24).

The second one is to compare the fit between the simulated and observed frequencies for the constraint variables in the geographical dataset. For each life zone, I compute the correlation between the values for the 47 constraint variables in the geographic aggregated dataset and in the simulated dataset. The distribution of these correlation coefficients is given in table 30. All coefficients, even the minimum, are particularly close to 1. It means that in every life zone, the synthetic values are extremely close to the original values for the constraint variables : synthetic populations are extremely similar to the original populations in terms of the defined socio-demographic and political characteristics. These results provides further evidence for the internal validity of this approach.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Correlation coefficient	1,632	1.000	0.0004	0.994	1.000	1.000	1.000	1.000

Table 30 – Distribution of the correlation coefficients between the observed and simulated frequencies of the constraint variables

### 5.b.4 External validation

To check the external validity of this approach, I compare the synthetic interpersonal trust variable to alternative variables measuring close substitutes or proxies to interpersonal trust. The absence of any correlation between these variables would shed serious doubts about the validity of the approach.

First of all, I compare the geographic distribution of the synthetic interpersonal trust variable and the map of the integration metric from [Le Bras and Todd \(2013\)](#). The integration metric is a 0-4 variable that summarizes 3 dimensions of social integration : family structures, type of settlement, and strength of the Church. The overlap between these two maps is striking : life zones with a higher synthetic interpersonal trust tend to be places that are well-integrated, in the West, the South-West, the Alps and Alsace. On the other hand, the regions around Paris, Bordeaux, and the Côte d’Azur display a lower level of interpersonal trust.

On top of that, the synthetic interpersonal trust variable is highly correlated to the revealed local trust index from [Algan et al. \(2019b\)](#) : the correlation coefficient is equal to 0.89 (see figures 11 and 12). This result is not surprising given the strong similarities in the way these variables were constructed<sup>12</sup>.

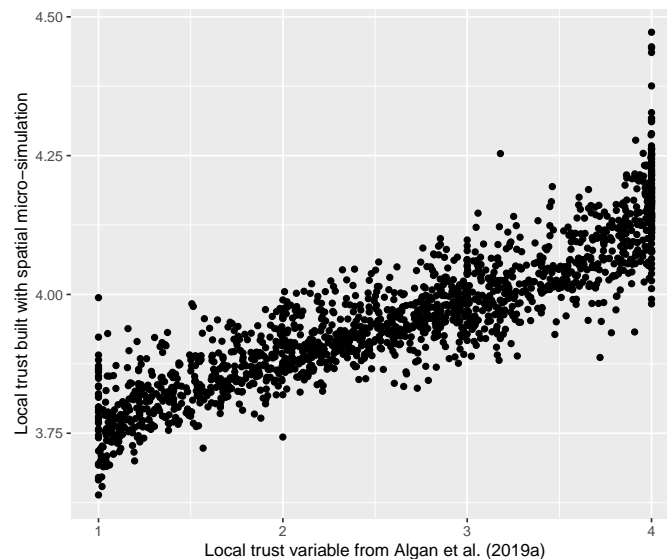
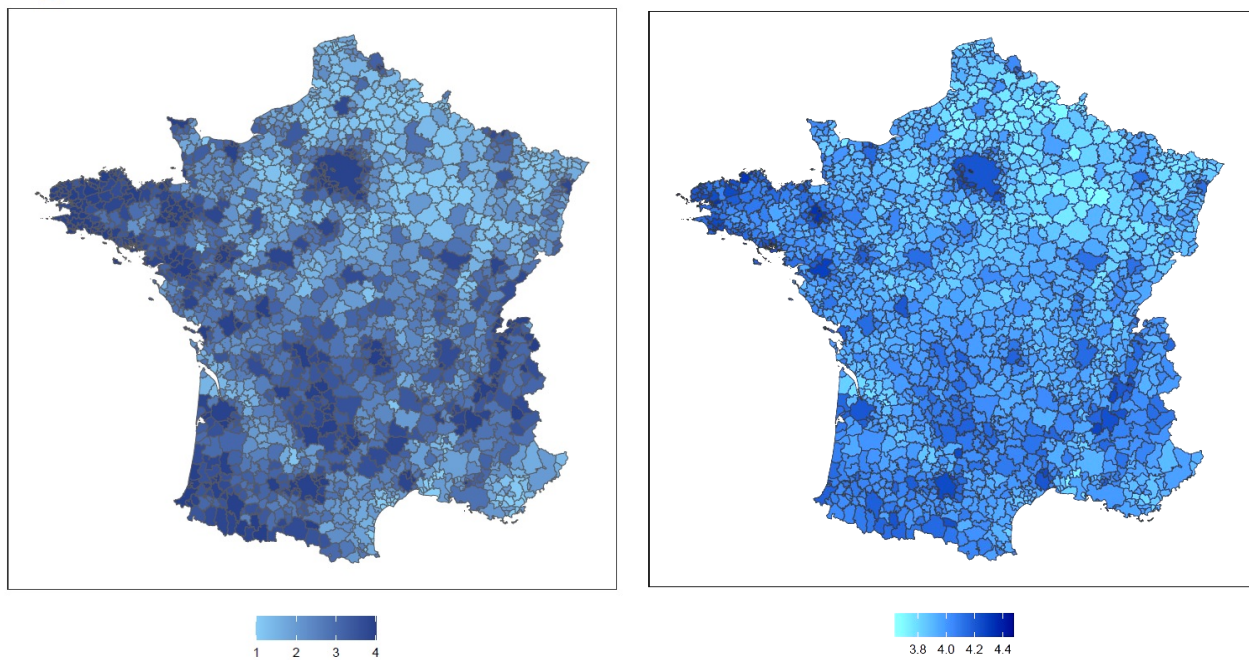


Figure 11 – Relationship between synthetic trust variables from [Algan et al. \(2019b\)](#) and the one built using spatial micro-simulation

---

<sup>12</sup>Figure 12a is built by averaging the municipal quartile value of the index of revealed trust from [Algan et al. \(2019b\)](#) at the life zone level, weighting by population. I thank Elisabeth Beasley and Daniel Cohen for access to the data.



(a) Synthetic interpersonal trust variable from [Algan et al. \(2019b\)](#) (b) Interpersonal trust variable built with spatial micro-simulation

Figure 12 – Maps of the synthetic trust variables at the locality level

From [Putnam \(1993, 2000\)](#), the density of associations has been considered a good indicator for the level of “social capital” of a given region. Using data from *Registre National des Associations* (National Register of Associations), in which each association is localized at the municipality level, I compute the average number of associations per 100 inhabitants at the life zone level. Data is missing for Alsace and Moselle, due to the distinct legal status of associations in this region. Discarding life zones from these regions, there is indeed a positive correlation (Person’s  $r$  equal to 0.22) between the local interpersonal trust synthetic variable and the density of associations (see figure 13).

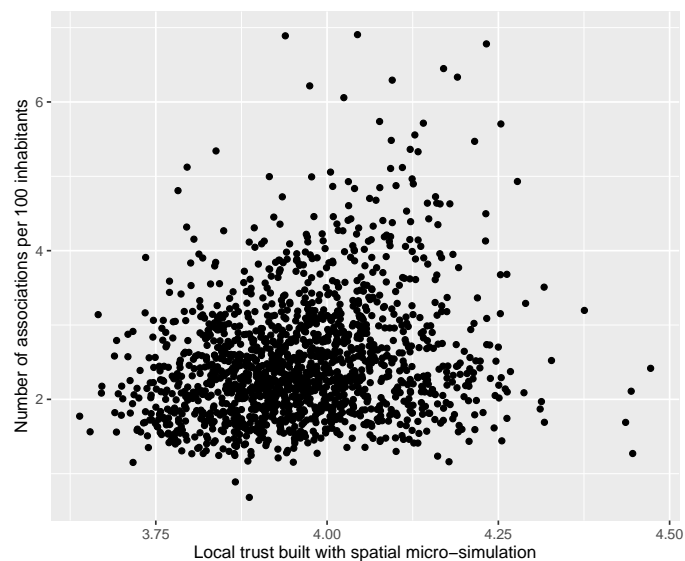


Figure 13 – Relationship between the synthetic interpersonal trust variable and the density of associations

### **5.b.5 Decomposition of the synthetic local interpersonal trust variable**

It is worth noting that, while the synthetic interpersonal trust variable is indeed a new metric measuring trust at the life zone level, it is nothing but a linear combination of the constraint variables used to create it. Indeed, as shown in table 31, a simple regression of the local interpersonal trust variable on the constraint variables is sufficient to drive the value of the R-squared to almost 1.

As a result, while the regression results will be interpreted in terms of interpersonal trust, we need to keep in mind that this variable is a linear combination of local socio-demographic and political variables. In section 2.g, the results of regressions on the variables that compose the local interpersonal trust variable are reported, and show that no single variable composing the interpersonal trust variable is actually driving the results.



	<i>Dependent variable:</i>
	Local synthetic interpersonal trust
Population age 20-39	-0.003*** (0.0002)
Population age 64 and above	0.002*** (0.0001)
Population with CAP-BEP (secondary vocational diploma)	0.0004** (0.0002)
Population with baccalauréat (high school diploma)	0.003*** (0.0003)
Population with higher education diploma	0.014*** (0.0001)
Macron vote in 2017 - First round	0.005*** (0.0003)
Le Pen vote in 2017 - First round	-0.006*** (0.0002)
Fillon vote in 2017 - First round	0.002*** (0.0002)
Mélenchon vote in 2017 - First round	0.006*** (0.0003)
Hamon vote in 2017 - First round	0.016*** (0.0005)
Abstention in 2017 - First round	-0.003*** (0.0002)
Constant	3.601*** (0.021)
Observations	1,632
R <sup>2</sup>	0.989

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 31 – Geographic decomposition of the interpersonal trust synthetic variable

### 5.b.6 Modalities of the linking variables

Gender (2 modalities) :

- Male
- Female

Age (5 modalities) :

- 20-29
- 30-39
- 40-49
- 50-64
- 65 and above

Education (4 modalities) :

- No diploma
- Vocational secondary diploma (CAP, BEP)
- Any high school diploma (baccalauréat général, professionnel, technologique)
- Higher education diploma

Vote in the first round of the 2017 Presidential election (7 modalities) :

- Abstention
- Vote for Fillon
- Vote for Hamon
- Vote for Macron
- Vote for Mélenchon
- Vote for Le Pen
- Other (vote for another candidate, blank or null vote)

### 5.b.7 Summary statistics tables

Linking variable	Share of the sample (in percent)
Female	52.7
Age 20-29	15.2
Age 30-39	15.5
Age 40-49	15.3
Age 50-64	26.7
Age 65 or above	27.3
No diploma	22.7
Vocational secondary school diploma (CAP, BEP)	27.8
ny high school diploma (baccalauréat général, professionnel, technologique)	19.1
Higher education diploma	30.4
Abstention in 2017	14.9
Fillon voter	11.7
Hamon voter	5.4
Macron voter	19.5
Mélenchon voter	18.0
Le Pen voter	15.9
Other voter (vote for another candidate, blank and null vote)	14.6

Table 32 – Summary statistics of the linking variables in the individual microdata

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Interpersonal trust	1,767	4.106	2.529	0	2	5	6	10

Table 33 – Summary statistics of the interpersonal trust variable in the individual microdata

s

## 5.c Building profiles of respondents using Latent Dirichlet Allocation

### 5.c.1 The LDA algorithm to build profiles of respondents to the short questionnaires

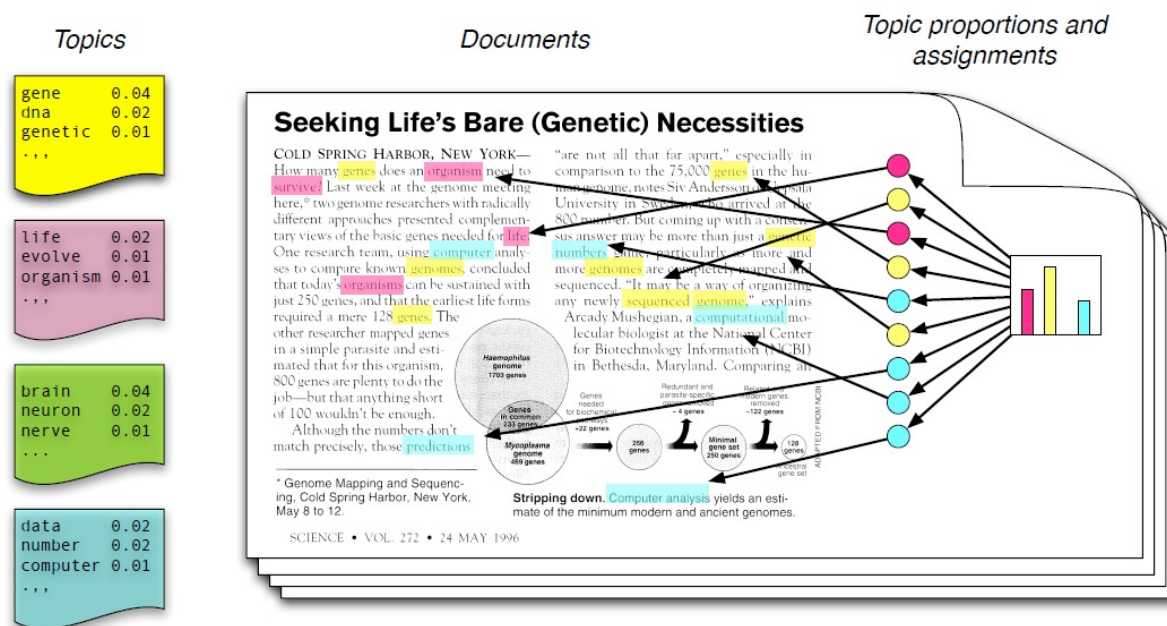


Figure 14 – Illustration of Latent Dirichlet Analysis from Blei (2012)

Figure 14 summarizes the construction and assumptions of LDA applied on text data. LDA posits that there exists a certain number of topics, which are probability distributions over a fixed vocabulary (left-hand side column). Certain topics (for instance, if the topic is nature) put a high probability for certain words (tree, river) and a low probability for others (computer, book). Each text is then characterized as a certain distribution of topics (for instance : 20% nature, 50% love, 30% happiness). The generation process for each text is as follows : first, randomly pick a topic from the topic distribution probabilities of this text, then randomly choose a word from this topic's probability distribution over the fixed vocabulary (see figure 14 from Blei (2012)). Note that the word order is not regarded here.

When it is run over a collection of texts, the LDA algorithm identifies two things :

- the topics, defined by their probability distributions over the fixed vocabulary.
- the probability distributions of topics for each text.

Note that the LDA algorithm does not provide any title for the topics that it defines. They have to be reconstructed by the researcher from the results of the analysis. For instance, if a given topic gives a big probability weight to words such as “bird”, “cat”, “dog”, then it can be inferred that this topic is about animals.

Instead of applying this algorithm to a collection of texts in order to study the underlying topics present in these texts, the contribution from Draca and Schwarz (2018) and Stantcheva (2019) is to apply it to the answers given by respondents to close-ended questions in survey data. Answers are treated as words, and topics are interpreted as ideologies. An ideology is thus defined as a probability distribution over the answers : if a given ideology gives a high probability for “No” to “Are you in favor of same-sex marriage ?” and “Yes” to “Do you believe abortion should be made harder ?”, then it is fairly easy to identify the considered ideology as conservative on social issues. Just like a text was originally defined in the traditional LDA

approach as the distribution of shares for each topic, each respondent here is modelled as a distribution of shares for several ideologies. What LDA does is that it defines these ideologies in a bottom-up approach, and identifies the shares that each ideology represents within each respondent.

In essence, LDA is used in this context to perform dimension reduction (or clustering) in our data of interest : here, the answers to GDN questionnaires. Respondents whose answers to this set of close-ended and open-ended questions differ across a very large number of dimensions are sorted into a small finite number of profiles. Alternative tools could have been used to perform dimension reduction, such as Part Component Analysis. However, [Draca and Schwarz \(2018\)](#) provide evidence that alternative clustering tools do not perform as well as Latent Dirichlet Analysis, which offers a flexible, micro-founded model that deals better with categorical data.

To build the profiles of respondents to the short questionnaires, I rely on the method described above, and use LDA to build a series of topics, with the answers to the short questionnaires treated as “words” and the list of answers as a “text” for the LDA algorithm. This is straightforward because all questions from the short questionnaires are close-ended questions. The topics produced by LDA are interpreted as profiles of respondents. I build four profiles of respondents using answers to the short questionnaires considered jointly.

Each respondent is then assigned to the profile whose inner share, computed by the algorithm, is the largest : if the LDA algorithm identifies that its probability distribution of ideologies is 20% “conservative” and 80% “liberal”, then it is sorted as a liberal. Eventually, I aggregate respondents per profile, at the life zone level, to compute the local share of each profile of respondents.

### **5.c.2 Construction of the profiles of respondents to the “Democracy and citizenship” long questionnaire**

To build the profiles of respondents to the long questionnaire, I draw inspiration from the method described above. For all questions of this questionnaire, I code answers to the questions into a finite set of possible answers. There are two cases : for the close-ended questions, I proceed like above, by relying on the proposed answers. For the open-ended questions, I explain below how I proceed. Then, once all answers are coded into this finite set of possible answers, I simply run the LDA algorithm exactly like in the previous section, to build two profiles of respondents, and then assign all respondents into either profile.

To code open-ended questions into a finite set of possible answers, I rely on a keywords-based approach to identify topics mentioned by respondents in their answers. This keyword-based approach performs as follows : whenever a certain keyword is used by the respondent in his/her answer, then the answer is coded as having mentioned this topic. Each answer can mention one, several, or no predefined topic. For instance, when asked “What do you think about the participation of citizens in elections and how to encourage them to participate more?”, if the respondent answers “It is a matter of education. I believe the vote should be made compulsory”. Then it matches two of our pre-defined topics : education, and compulsory vote.

The list of keywords for each question is documented in the appendix table 24. These keywords are identified by using the synthesis produced by OpinionWay, a polling company, and published on the GDN’s webpage. OpinionWay partnered with QWAM, an artificial intelligence company, who used its own natural language processing tools called QWAC text analytics, to identify the topics that were mentioned in the answers. [Bellet et al. \(2020\)](#) highlight that their methodology is opaque and thus cannot be reproduced : the authors try to reproduce their topic modelling results using state-of-the-art natural language processing techniques, but fail to get results that are close to theirs. The list of keywords is custom-made and aims to cover the most-used topics highlighted by QWAC. For instance, QWAC identifies the environment as a topic that was much mentioned to the question “What are the civic behaviors that should be promoted

in our daily or collective life?”. To identify respondents who mentioned this topic, I use the keywords “environnement”, “déchets”, “pollution”. Only the topics that gathered a significant number of answers are kept for the analysis. Similarly, questions whose answers are so disparate that no topic was mentioned by a sizable portion of the respondents are excluded from the analysis. In any case, this analysis is mostly robust to the exclusion of topics mentioned by a tiny portion of respondents : LDA, when applied on text data, is mostly driven by words which occur relatively often.

This keyword-based approach is arguably quite raw. It does not take into account negations (a respondent who wrote “I am not in favor of proportional representation” will be treated just like a respondent who wrote “I want more proportional representation” : both mentioned the topic “proportional representation”). Therefore, I took great care, when selecting the questions included in this analysis and selecting the topics and corresponding keywords, to design them so that they are subject to the minimum ambiguity. The main advantage of this approach is its great flexibility. It is also computationally efficient, which makes it a very attractive strategy given the amount of data to be processed. On top of that, keywords approach performs well when it is about identifying simple answers, for instance a type of elected representative (“which kind of elected representative should see their number decrease ?”). It is less appropriate to use it when respondents are asked about complex ideas, which we did not attempt to do here (a reason, a personal experience, a full-fledged proposal...).

Note that alternative natural language processing methods to classify responses to open-ended questions into topics could have been used instead of the keywords approach. They are reviewed in [Bellet et al. \(2020\)](#). These methods offer a better coverage of the topics since they can identify complex thoughts and reasoning within answers. However, they were too computationally intensive to be included in this thesis, but they offer attractive ways ahead for further research.

### 5.c.3 Profiles built from LDA analysis of the short questionnaires : top 15 answers per profile

#### Profile 1

- Non, à “pensez-vous que les taxes sur le diesel et sur l’essence peuvent permettre de modifier les comportements des utilisateurs ?”
- Non, à “seriez-vous prêt à payer un impôt pour encourager des comportements bénéfiques à la collectivité comme la fiscalité écologique ou la fiscalité sur le tabac ou l’alcool ?”
- Les dépenses de l’Etat, à “afin de baisser les impôts et réduire la dette, quelles dépenses publiques faut-il réduire en priorité ?”
- Physiques, à “pour accéder à certains services publics, vous avez avant tout des besoins...”
- Physique, à “si vous rencontrez des difficultés pour effectuer vos démarches administratives sur Internet, de quel accompagnement souhaiteriez-vous bénéficier ?”
- Oui, à “avez-vous déjà renoncé à des droits / des allocations en raison de démarches administratives trop complexes ?”
- Entreprises, à “et qui doit être en priorité concerné par le financement de la transition écologique ?”
- Commune, à “Quels sont les niveaux de collectivités territoriales auxquels vous êtes le plus attaché ?”
- Jusqu’à 5km, à “lorsqu’un déplacement est nécessaire pour effectuer une démarche administrative, quelle distance pouvez-vous parcourir sans difficulté ?”
- Oui, à “faut-il avoir davantage recours au référendum au niveau national ?”
- Département, à “Quels sont les niveaux de collectivités territoriales auxquels vous êtes le plus attaché ?”
- Réduire la dépense publique, à “afin de réduire le déficit public de la France qui dépense plus qu’elle ne gagne, pensez-vous qu’il faut avant tout: ...”

- Non, à “pensez-vous que vos actions en faveur de l’environnement peuvent vous permettre de faire des économies ?”
- Oui, à “faut-il avoir davantage recours au référendum au niveau local ?”
- Oui, à “pensez-vous qu’il serait souhaitable de réduire le nombre de parlementaires (députés + sénateurs = 925) ?”

#### Profil 2 :

- Oui, à “pensez-vous que vos actions en faveur de l’environnement peuvent vous permettre de faire des économies ?”
- Tout le monde, à “et qui doit être en priorité concerné par le financement de la transition écologique ?”
- Par la fiscalité écologique et par le budget général de l’Etat, à “selon vous, la transition écologique doit être avant tout financée : ...”
- Oui, à “pensez-vous que les taxes sur le diesel et sur l’essence peuvent permettre de modifier les comportements des utilisateurs ?”
- Oui, à “seriez-vous prêt à payer un impôt pour encourager des comportements bénéfiques à la collectivité comme la fiscalité écologique ou la fiscalité sur le tabac ou l’alcool ?”
- Non, à “diriez-vous que vous connaissez les aides et dispositifs qui sont aujourd’hui proposés par l’Etat, les collectivités, les entreprises et les associations pour l’isolation et le chauffage des logements, et pour les déplacements ?”
- A financer des investissements en faveur du climat et à financer des aides pour accompagner les Français dans la transition écologique, à “À quoi les recettes liées aux taxes sur le diesel et l’essence doivent-elles avant tout servir ?”
- Satisfaisante, à “diriez-vous que l’application de la laïcité en France est aujourd’hui: ...”
- Non, à “faut-il avoir davantage recours au référendum au niveau national ?”
- Oui, à “seriez-vous d’accord pour qu’un agent public effectue certaines démarches à votre place ?”
- Bonne chose, à “Que pensez-vous des services publics itinérants (bus de services publics) ?”
- Bonne chose, à “Que pensez-vous des agents publics polyvalents susceptibles de vous accompagner dans l’accomplissement de plusieurs démarches quelle que soit l’administration concernée ?”
- Intercommunalité, à “Quels sont les niveaux de collectivités territoriales auxquels vous êtes le plus attaché ?”
- Numériques, à “pour accéder à certains services publics, vous avez avant tout des besoins...”

#### Profil 3 :

- Régionales, départementales et législatives, à “selon vous, faut-il introduire une dose de proportionnelle pour certaines élections, lesquelles ?”
- Oui, à “faut-il avoir davantage recours au référendum au niveau local ?”
- Oui, à “pensez-vous qu’il serait souhaitable de réduire le nombre de parlementaires (députés + sénateurs = 925) ?”
- Oui, à “faut-il tirer au sort des citoyens non élus pour les associer à la décision publique ?”
- Oui, à “faut-il rendre le vote obligatoire ?”
- A améliorer, à “diriez-vous que l’application de la laïcité en France est aujourd’hui:”
- Oui, à “pensez-vous qu’il y a trop d’échelons administratifs en France ?”
- Oui, à “seriez-vous d’accord pour qu’un agent public effectue certaines démarches à votre place ?”
- Bonne chose, à “Que pensez-vous des agents publics polyvalents susceptibles de vous accompagner dans l’accomplissement de plusieurs démarches quelle que soit l’administration concernée ?”
- Bonne chose, à “Que pensez-vous des services publics itinérants (bus de services publics) ?”

- Bonne chose, à “Que pensez-vous du regroupement dans un même lieu de plusieurs services publics (Maisons de services au public) ?”
- Oui, à “faut-il avoir davantage recours au référendum au niveau national ?”
- Bonne chose, à “Que pensez-vous du service public sur prise de rendez-vous ?”

#### Profile 4 :

- Numériques, à “pour accéder à certains services publics, vous avez avant tout des besoins...”
- Bonne chose, à “Que pensez-vous du service public sur prise de rendez-vous ?”
- Oui, à “pensez-vous qu’il y a trop d’échelons administratifs en France ?”
- Bonne chose, à “Que pensez-vous du regroupement dans un même lieu de plusieurs services publics (Maisons de services au public) ?”
- Bonne chose, à “Que pensez-vous des agents publics polyvalents susceptibles de vous accompagner dans l’accomplissement de plusieurs démarches quelle que soit l’administration concernée ?”
- Non, à “avez-vous déjà renoncé à des droits / des allocations en raison de démarches administratives trop complexes ?”
- Bonne chose, à “Que pensez-vous des services publics itinérants (bus de services publics) ?”
- Oui, à “savez-vous quels sont les différents échelons administratifs (Etat, collectivités territoriales comme la région, la commune, opérateurs comme par exemple Pole Emploi ou la CAF) qui gèrent les différents services publics dans votre territoire ?”
- Réduire la dépense publique, à “afin de réduire le déficit public de la France qui dépense plus qu’elle ne gagne, pensez-vous qu’il faut avant tout: ...”
- Téléphonique, à “si vous rencontrez des difficultés pour effectuer vos démarches administratives sur Internet, de quel accompagnement souhaiteriez-vous bénéficier ?”
- Non, à “faut-il tirer au sort des citoyens non élus pour les associer à la décision publique ?”
- Non, à “seriez-vous d’accord pour qu’un agent public effectue certaines démarches à votre place ?”
- Oui, à “pensez-vous qu’il serait souhaitable de réduire le nombre de parlementaires (députés + sénateurs = 925) ?”
- Non, à “faut-il avoir davantage recours au référendum au niveau national ?”
- Il ne faut pas introduire de proportionnelle, à “selon vous, faut-il introduire une dose de proportionnelle pour certaines élections, lesquelles ?”



#### 5.c.4 Profiles built from LDA analysis of the “Democracy and Citizenship” long questionnaire : top 15 answers per profile

Profile 1 :

- Oui, à “pensez-vous qu’il faille instaurer des contreparties aux différentes allocations de solidarité ?”
- Oui, à “pensez-vous qu’il serait souhaitable de réduire le nombre d’élus (hors députés et sénateurs) ?”
- Non, à “en dehors des élus politiques, faut-il donner un rôle plus important aux associations et aux organisations syndicales et professionnelles ?”
- Oui, à “en matière d’immigration, une fois nos obligations d’asile remplies, souhaitez-vous que nous puissions nous fixer des objectifs annuels définis par le Parlement ?”
- Oui, à “faut-il les (le Sénat et le Conseil économique, social et environnemental) transformer ?”
- Non, à “faut-il faciliter le déclenchement du référendum d’initiative partagée ?”
- Elus départementaux, régionaux, et municipaux, à “de quels élus pensez-vous qu’il serait souhaitable de réduire le nombre (hors députés et sénateurs) ?”
- Non, à “faut-il prendre en compte le vote blanc ?”
- Proportionnelle, à “Que faudrait-il faire pour mieux représenter les différentes sensibilités politiques ?”
- Député, Maire, à “en qui faites-vous le plus confiance pour vous faire représenter dans la société et pourquoi ?”
- Respect, à “Quels sont les comportements civiques qu’il faut promouvoir dans notre vie quotidienne ou collective ?”
- Sanction, à “Que peuvent et doivent faire les pouvoirs publics pour répondre aux incivilités ?”

Profile 2 :

- Oui, à “en dehors des élus politiques, faut-il donner un rôle plus important aux associations et aux organisations syndicales et professionnelles ?”
- Association, à “a quel type d’associations ou d’organisations ? Et avec quel rôle ?”
- Non, à “pensez-vous qu’il serait souhaitable de réduire le nombre d’élus (hors députés et sénateurs) ?”
- Oui, à “faut-il prendre en compte le vote blanc ?”
- Non, à “pensez-vous qu’il faille instaurer des contreparties aux différentes allocations de solidarité ?”
- Une bonne chose, à “le non-cumul des mandats instauré en 2017 pour les parlementaires (députés et sénateurs) est : ...”
- Ne sais pas, à “faut-il faciliter le déclenchement du référendum d’initiative partagée ?”
- Solidarité, à “Quels sont les comportements civiques qu’il faut promouvoir dans notre vie quotidienne ou collective ?”
- Les élus en général, à “en qui faites-vous le plus confiance pour vous faire représenter dans la société et pourquoi ?”
- Syndicats, à “a quel type d’associations ou d’organisations ? Et avec quel rôle ?”
- Respect, à “Quels sont les comportements civiques qu’il faut promouvoir dans notre vie quotidienne ou collective ?”
- Sexisme, à “Quelles sont les discriminations les plus répandues dont vous êtes témoin ou victime ?”
- Ecole et travail, à “Quelles sont, selon vous, les modalités d’intégration les plus efficaces et les plus justes à mettre en place aujourd’hui dans la société ?”
- Non, à “en matière d’immigration, une fois nos obligations d’asile remplies, souhaitez-vous que nous puissions nous fixer des objectifs annuels définis par le Parlement ?”

## 5.d Conditional scatter plots for the geographic cross-sectional analysis

This section presents some graphical representation of the cross-sectional relationship between local interpersonal trust and the GDN variables of interest, by plotting the conditional scatter plot and linear trend line between local interpersonal trust and the variable of interest, conditioning on *département* fixed effects and the log-transformation of population density. In effect, these graphs plot the residuals of the regressions of local interpersonal trust variable and of the variable of interest on *département* fixed effects and the log-transformation of population density.

### 5.d.1 Participation variables

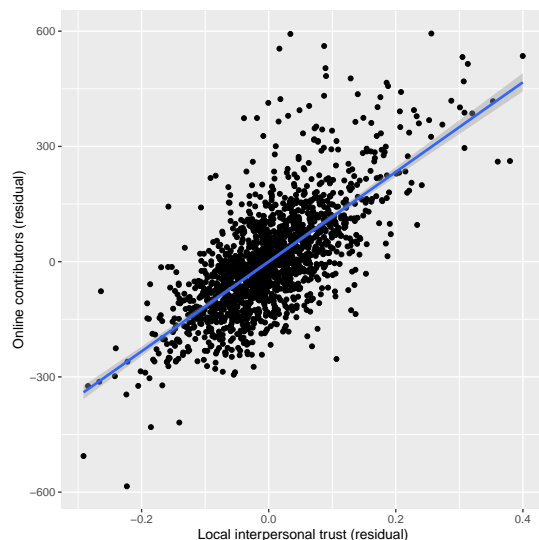
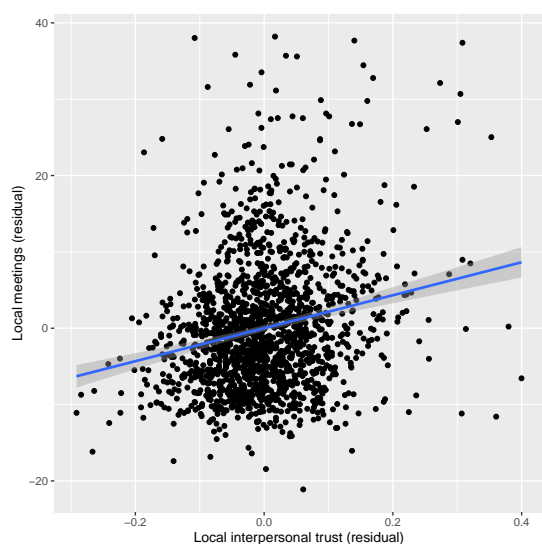
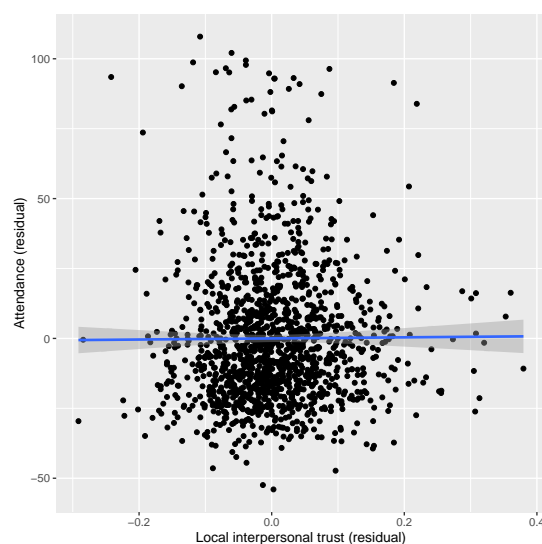


Figure 15 – Conditional scatter plot for extensive margin online participation variable of interest : share of online contributors

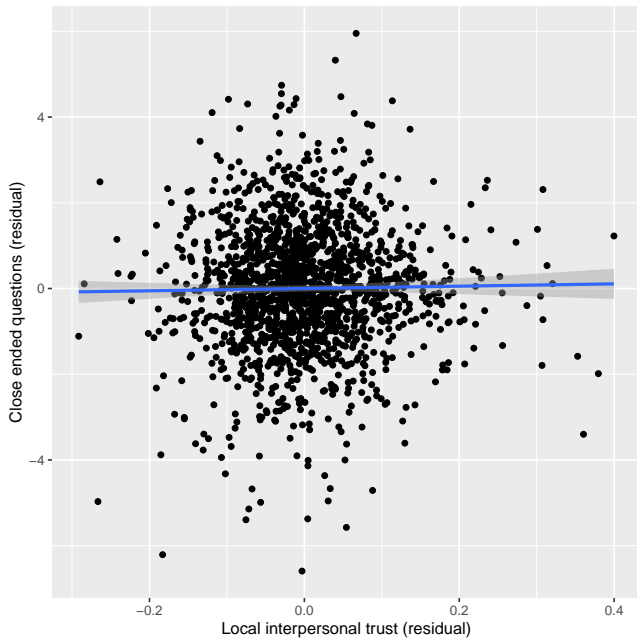


(a) Local meetings

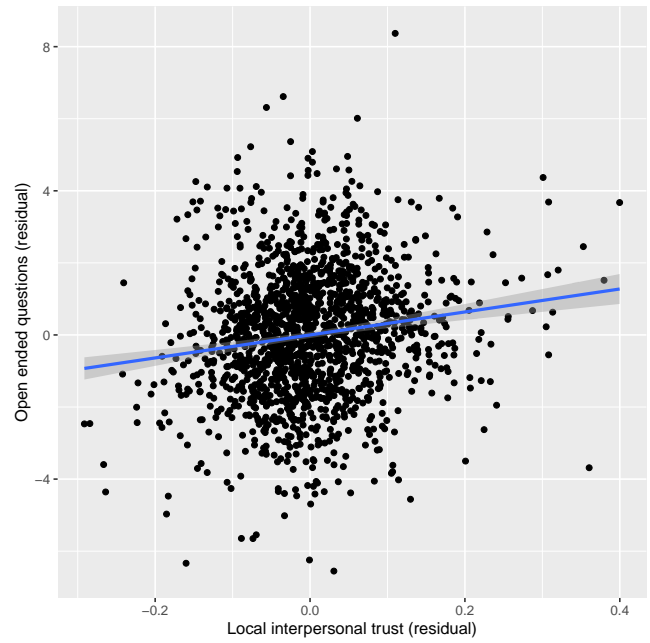


(b) Attendance

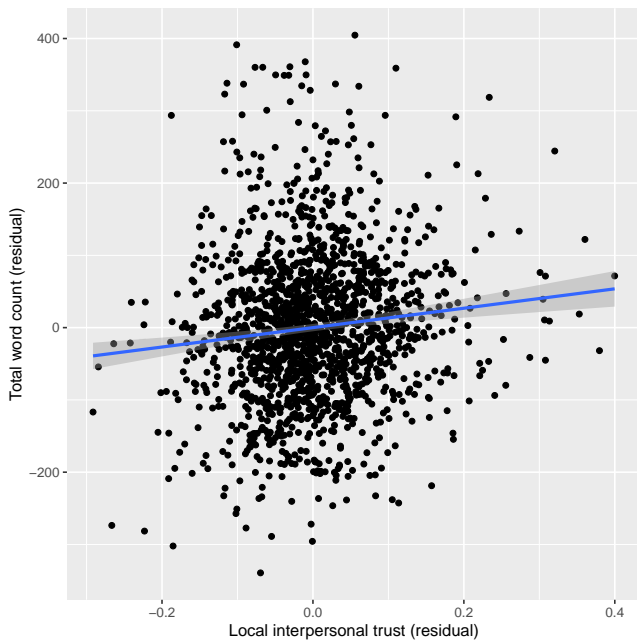
Figure 16 – Conditional scatter plots for offline participation variables of interest



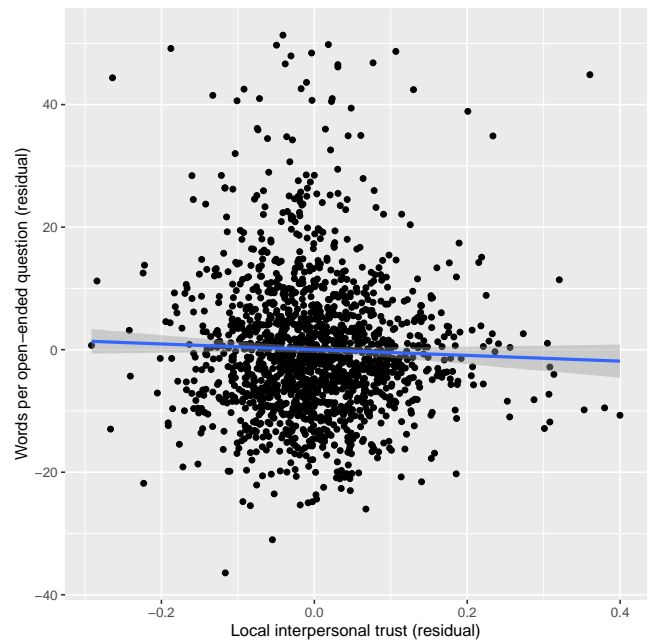
(a) Close-ended questions



(b) Open-ended questions



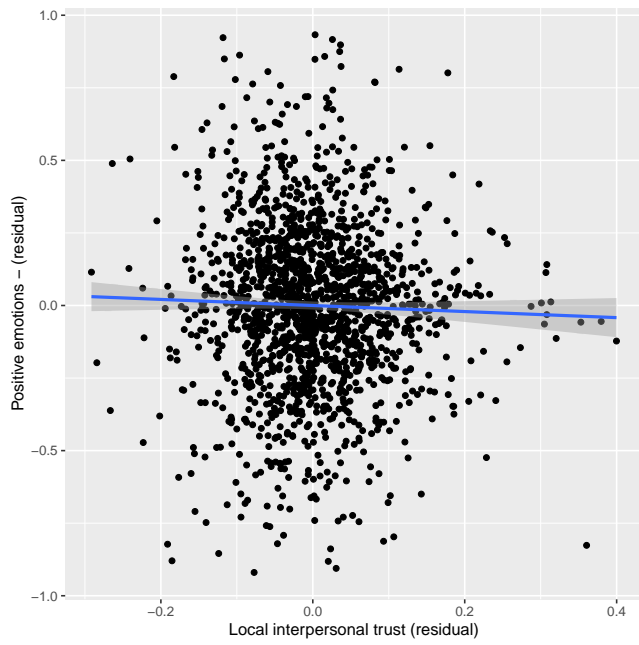
(c) Total word count



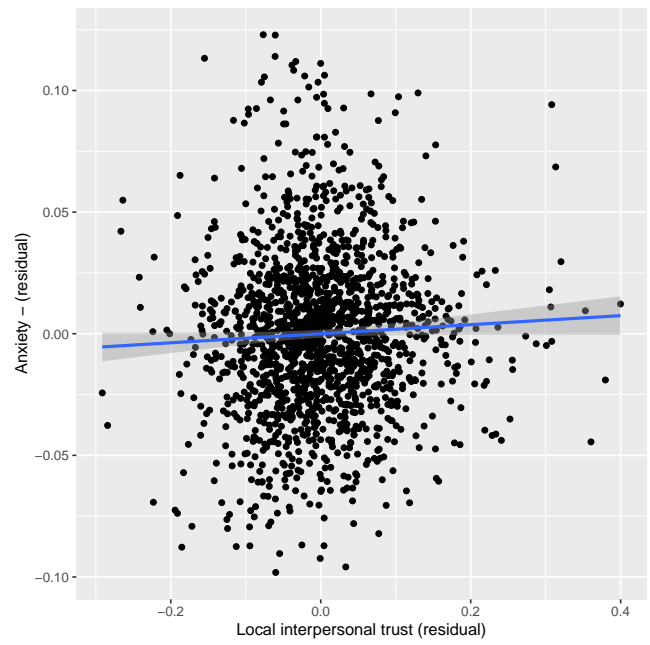
(d) Words per open-ended question

Figure 17 – Conditional scatter plots for intensive margin online participation variables of interest

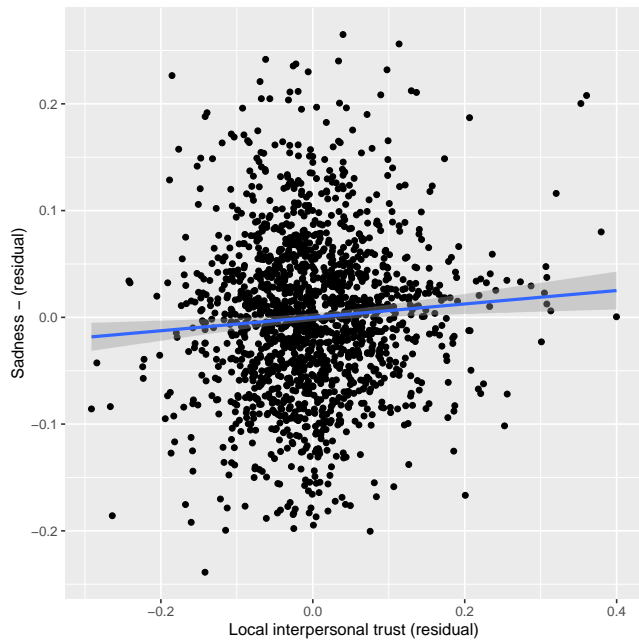
## 5.d.2 Lexical variables



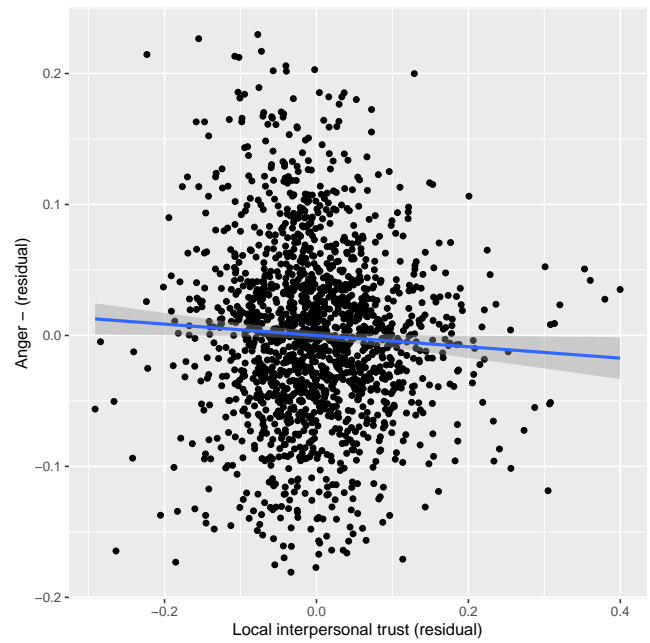
(a) Positive emotions



(b) Anxiety

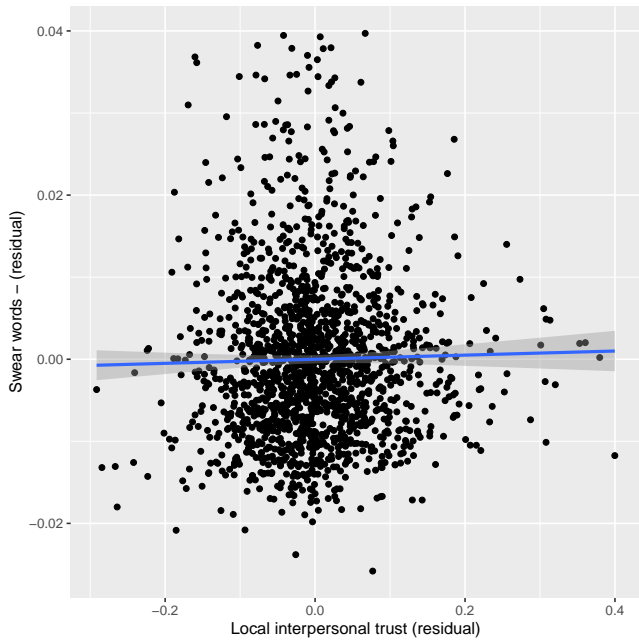


(c) Sadness

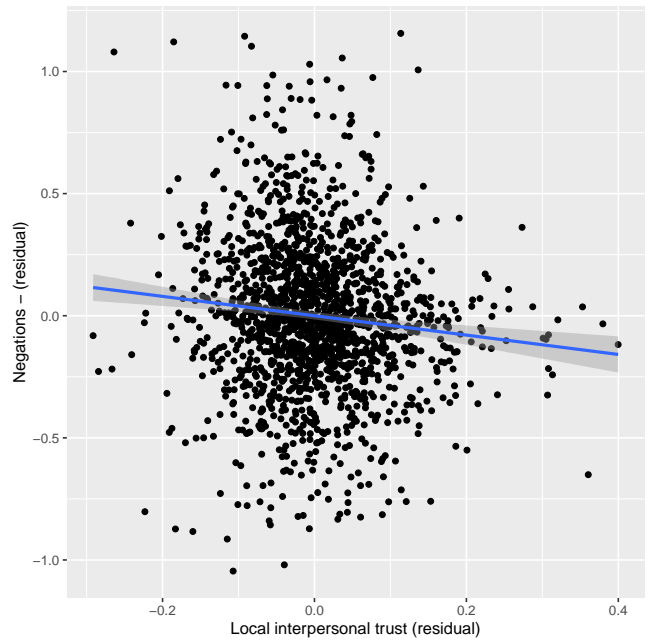


(d) Anger

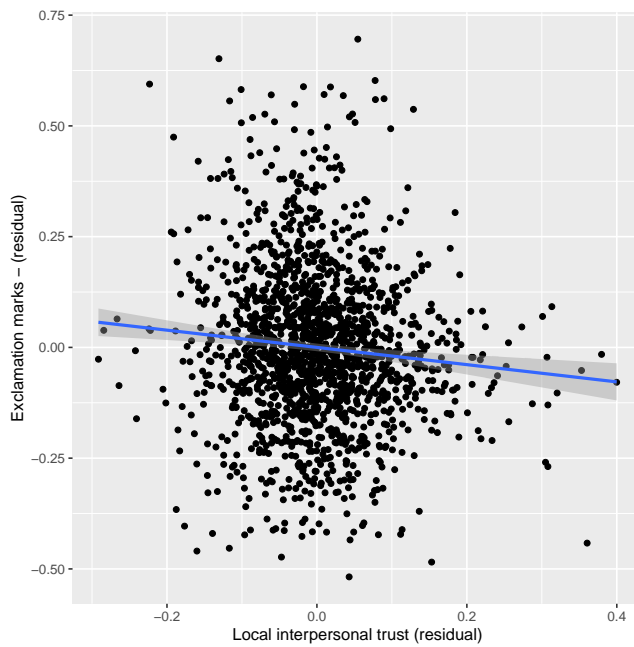
Figure 18 – Conditional scatter plots for dictionary-based lexical variables of interest (1/2)



(a) Swear words



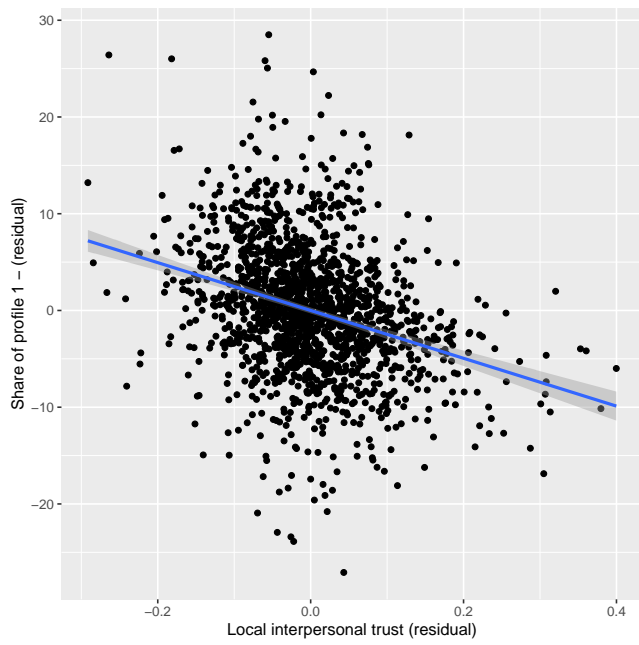
(b) Negations



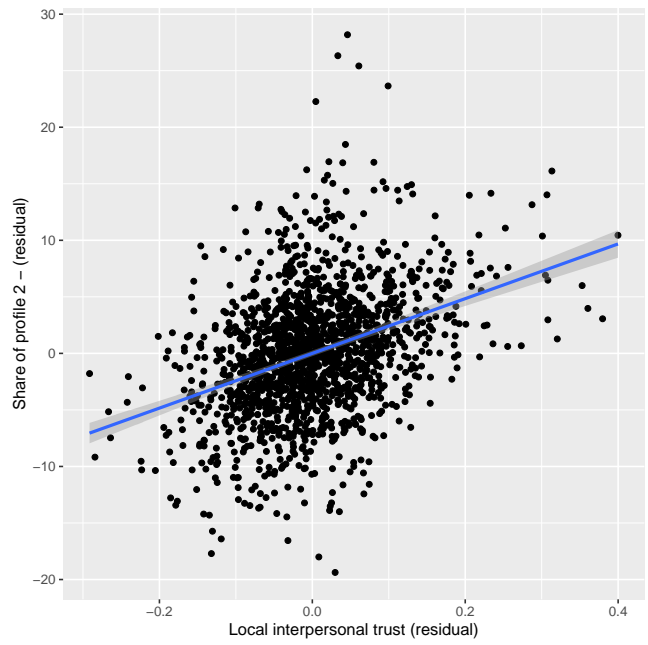
(c) Exclamation marks

Figure 19 – Conditional scatter plots for dictionary-based lexical variables of interest (2/2)

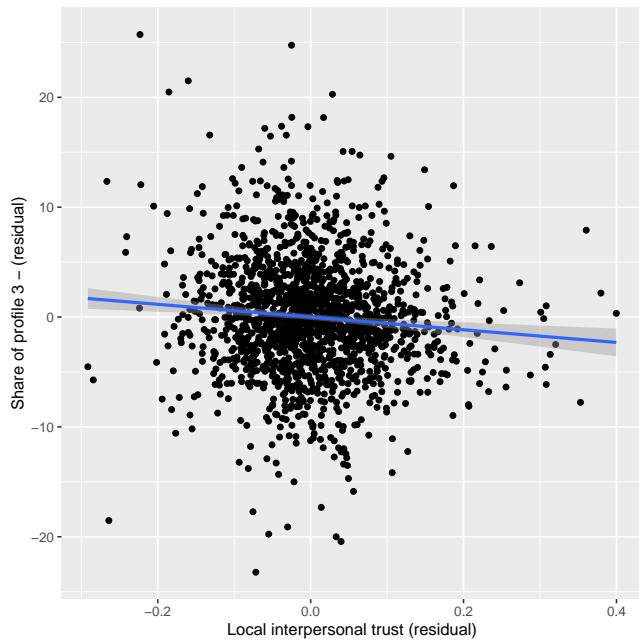
### 5.d.3 Profiles of respondents variables



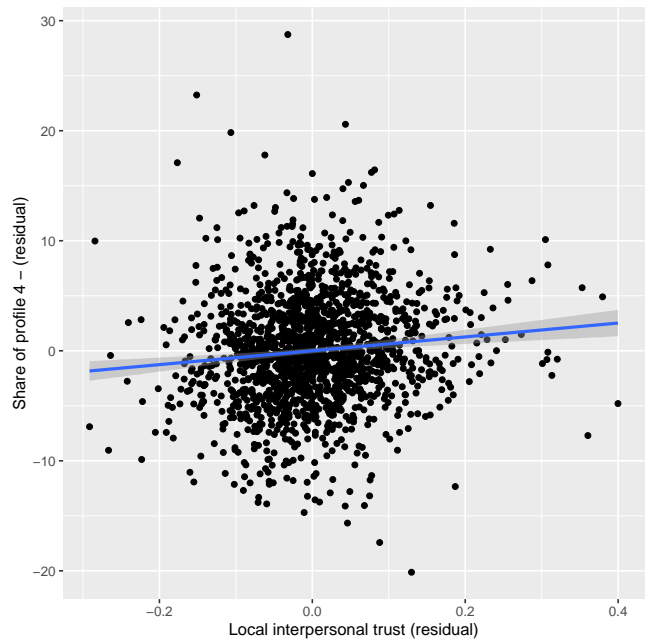
(a) Profile 1



(b) Profile 2



(c) Profile 3



(d) Profile 4

Figure 20 – Conditional scatter plots for the profiles of respondents to the short questionnaires

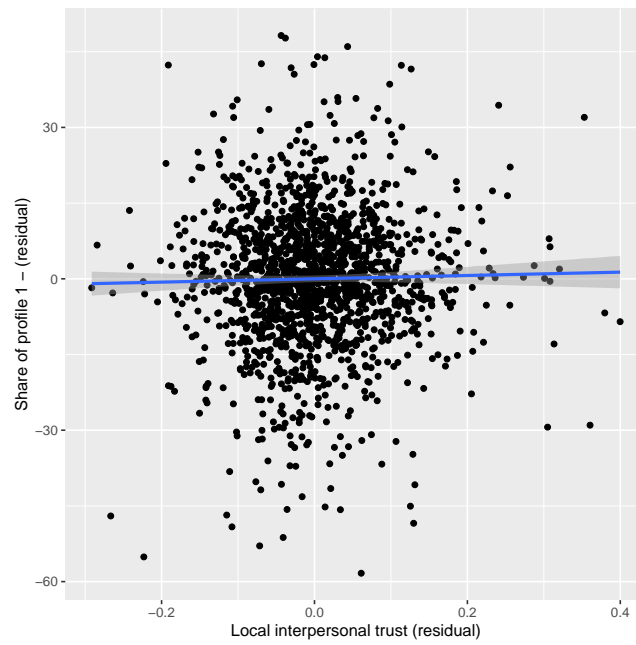


Figure 21 – Conditional scatter plots for the profile of respondents to the “Democracy and citizenship” long questionnaire

### 5.e Robustness check - Regression results with disaggregated local interpersonal trust variable

	<i>Dependent variable:</i>				
	Online contributors	Close-ended questions	Open-ended questions	Total word count	Words per open-ended question
	(1)	(2)	(3)	(4)	(5)
Population age 20-39	-5.544*** (1.648)	0.031 (0.028)	-0.080** (0.033)	-2.633 (1.969)	0.183 (0.217)
Population age 64 and above	9.232*** (1.070)	-0.014 (0.018)	0.020 (0.022)	1.260 (1.279)	0.030 (0.141)
Population with CAP-BEP (vocational diploma)	-0.421 (2.076)	-0.018 (0.035)	0.014 (0.042)	-1.768 (2.481)	-0.170 (0.273)
Population with baccalaureat	12.034*** (3.044)	0.053 (0.052)	0.080 (0.061)	-2.686 (3.637)	-0.876** (0.400)
Population with higher education diploma	32.137*** (1.485)	0.037 (0.025)	0.110*** (0.030)	4.233** (1.774)	-0.262 (0.195)
Macron vote in 2017 - First round	13.792*** (3.000)	0.035 (0.051)	-0.053 (0.060)	-5.959* (3.585)	-0.462 (0.395)
Le Pen vote in 2017 - First round	15.924*** (2.500)	0.064 (0.043)	-0.020 (0.050)	-0.847 (2.988)	-0.143 (0.329)
Fillon vote in 2017 - First round	11.935*** (2.520)	0.043 (0.043)	-0.040 (0.051)	0.738 (3.011)	0.272 (0.331)
Mélenchon vote in 2017 - First round	14.412*** (2.622)	0.059 (0.045)	0.051 (0.053)	5.143 (3.133)	0.094 (0.345)
Hamon vote in 2017 - First round	10.298** (4.770)	-0.086 (0.081)	-0.110 (0.096)	-2.976 (5.700)	0.537 (0.627)
Abstention in 2017 - First round	6.348*** (2.383)	0.026 (0.041)	0.003 (0.048)	-1.647 (2.848)	-0.344 (0.313)
Log-transformation of population density	22.716*** (4.707)	-0.024 (0.080)	-0.072 (0.095)	4.276 (5.624)	0.785 (0.619)
Constant	-1,263.617*** (220.369)	18.940*** (3.752)	9.768** (4.444)	652.213** (263.327)	78.768*** (28.984)
Département fixed effects	X	X	X	X	X
Observations	1,632	1,632	1,632	1,632	1,632
R <sup>2</sup>	0.803	0.090	0.147	0.103	0.084

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 34 – Cross-sectional relationship between socio-demographic and political variables and online participation to Grand Débat National



	<i>Dependent variable:</i>	
	Local meetings	Attendance
	(1)	(2)
Population age 20-39	-0.278* (0.159)	-0.712 (0.509)
Population age 64 and above	0.172* (0.103)	0.483 (0.344)
Population with CAP-BEP (vocational diploma)	-0.234 (0.201)	-0.632 (0.646)
Population with baccalaureat	0.236 (0.294)	0.615 (0.965)
Population with higher education diploma	0.011 (0.144)	0.602 (0.469)
Macron vote in 2017 - First round	0.327 (0.290)	0.487 (0.923)
Le Pen vote in 2017 - First round	0.077 (0.242)	0.899 (0.759)
Fillon vote in 2017 - First round	0.242 (0.244)	0.279 (0.784)
Mélenchon vote in 2017 - First round	0.494* (0.253)	-0.015 (0.804)
Hamon vote in 2017 - First round	0.629 (0.461)	-0.100 (1.581)
Abstention in 2017 - First round	-0.157 (0.230)	1.552** (0.739)
Log-transformation of population density	-1.325*** (0.455)	6.013*** (1.473)
Constant	-0.469 (21.298)	-19.232 (67.388)
Département fixed effects	X	X
Observations	1,632	1,396
R <sup>2</sup>	0.177	0.154

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 35 – Cross-sectional relationship between socio-demographic and political variables and offline participation to Grand Débat National

	<i>Dependent variable:</i>						
	Positive emotions	Anxiety	Sadness	Anger	Swear words	Negations	Exclamation marks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Population age 20-39	0.002 (0.005)	0.001 (0.001)	-0.002 (0.001)	-0.002 (0.001)	0.0001 (0.0002)	-0.003 (0.006)	-0.004 (0.003)
Population age 64 and above	0.004 (0.004)	0.001* (0.0004)	0.0003 (0.001)	-0.001 (0.001)	0.0002 (0.0001)	-0.001 (0.004)	-0.003 (0.002)
Population with CAP-BEP (vocational diploma)	-0.012* (0.007)	-0.002*** (0.001)	-0.003 (0.002)	0.001 (0.002)	-0.00004 (0.0002)	-0.015** (0.007)	0.005 (0.004)
Population with baccalaureat	-0.010 (0.010)	0.002* (0.001)	0.0002 (0.003)	-0.002 (0.002)	0.001*** (0.0004)	0.010 (0.011)	-0.001 (0.006)
Population with higher education diploma	-0.004 (0.005)	-0.001 (0.001)	0.001 (0.001)	-0.0002 (0.001)	-0.0002 (0.0002)	-0.019*** (0.005)	-0.002 (0.003)
Macron vote in 2017 - First round	0.010 (0.010)	0.001 (0.001)	-0.003 (0.003)	0.001 (0.002)	0.0002 (0.0004)	0.027** (0.011)	-0.009 (0.006)
Le Pen vote in 2017 - First round	0.001 (0.008)	0.001 (0.001)	-0.001 (0.002)	0.002 (0.002)	-0.00004 (0.0003)	0.021** (0.009)	-0.007 (0.005)
Fillon vote in 2017 - First round	-0.011 (0.008)	0.0005 (0.001)	-0.001 (0.002)	0.001 (0.002)	-0.0001 (0.0003)	0.015* (0.009)	-0.004 (0.005)
Mélenchon vote in 2017 - First round	-0.003 (0.009)	0.001 (0.001)	-0.001 (0.002)	-0.0001 (0.002)	-0.0001 (0.0003)	0.019** (0.009)	-0.003 (0.005)
Hamon vote in 2017 - First round	-0.022 (0.016)	-0.001 (0.002)	0.001 (0.004)	0.002 (0.004)	-0.0002 (0.001)	0.002 (0.017)	-0.017* (0.010)
Abstention in 2017 - First round	0.0002 (0.008)	-0.0001 (0.001)	-0.001 (0.002)	0.001 (0.002)	0.0001 (0.0003)	0.017** (0.009)	-0.007 (0.005)
Log-transformation of population density	0.021 (0.015)	-0.0005 (0.002)	-0.004 (0.004)	-0.001 (0.004)	0.0004 (0.001)	-0.025 (0.017)	-0.008 (0.010)
Constant	3.417*** (0.724)	0.063 (0.085)	0.592*** (0.191)	0.222 (0.173)	-0.006 (0.026)	1.456* (0.796)	1.187*** (0.451)
Département fixed effects	X	X	X	X	X	X	X
Observations	1,632	1,632	1,632	1,632	1,632	1,632	1,632
R <sup>2</sup>	0.074	0.073	0.076	0.090	0.098	0.109	0.105

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 36 – Cross-sectional relationship between socio-demographic and political variables and dictionary-based lexical variables of interest

	<i>Dependent variable:</i>		
	Share of six-letter words	Mean word syllables	Mean sentence length
	(1)	(2)	(3)
Population age 20-39	-0.030* (0.017)	-0.001** (0.001)	0.078** (0.031)
Population age 64 and above	-0.010 (0.011)	-0.001** (0.0003)	0.059*** (0.020)
Population with CAP-BEP (vocational diploma)	0.047** (0.022)	-0.001 (0.001)	-0.038 (0.040)
Population with baccalaureat	-0.004 (0.032)	0.002 (0.001)	0.002 (0.058)
Population with higher education diploma	0.073*** (0.016)	0.001* (0.0005)	-0.032 (0.028)
Macron vote in 2017 - First round	-0.022 (0.032)	-0.002* (0.001)	0.143** (0.057)
Le Pen vote in 2017 - First round	0.006 (0.026)	-0.001 (0.001)	0.072 (0.048)
Fillon vote in 2017 - First round	0.020 (0.027)	-0.001 (0.001)	0.057 (0.048)
Mélenchon vote in 2017 - First round	0.032 (0.028)	-0.001 (0.001)	0.120** (0.050)
Hamon vote in 2017 - First round	0.052 (0.050)	0.0003 (0.002)	-0.019 (0.091)
Abstention in 2017 - First round	0.023 (0.025)	-0.001 (0.001)	0.008 (0.045)
Log-transformation of population density	0.103** (0.050)	-0.002 (0.002)	-0.118 (0.090)
Constant	28.572*** (2.326)	1.916*** (0.071)	9.874** (4.199)
Département fixed effects	X	X	X
Observations	1,632	1,632	1,632
R <sup>2</sup>	0.151	0.107	0.101

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 37 – Cross-sectional relationship between socio-demographic and political variables and lexical complexity variables

	<i>Dependent variable:</i>			
	Profile 1	Profile 2	Profile 3	Profile 4
	(1)	(2)	(3)	(4)
Population age 20-39	0.125 (0.121)	0.103 (0.097)	-0.006 (0.101)	-0.223** (0.095)
Population age 64 and above	-0.022 (0.078)	-0.007 (0.063)	-0.002 (0.066)	0.031 (0.061)
Population with CAP-BEP (vocational diploma)	0.090 (0.152)	0.051 (0.122)	-0.092 (0.128)	-0.050 (0.119)
Population with baccalaureat	-0.155 (0.223)	-0.277 (0.179)	0.031 (0.187)	0.400** (0.175)
Population with higher education diploma	-0.437*** (0.109)	0.306*** (0.087)	-0.082 (0.091)	0.213** (0.085)
Macron vote in 2017 - First round	-0.150 (0.219)	-0.148 (0.176)	0.105 (0.185)	0.193 (0.172)
Le Pen vote in 2017 - First round	0.018 (0.183)	-0.396*** (0.147)	0.094 (0.154)	0.284** (0.143)
Fillon vote in 2017 - First round	-0.176 (0.184)	0.005 (0.148)	-0.092 (0.155)	0.263* (0.145)
Mélenchon vote in 2017 - First round	-0.008 (0.192)	-0.179 (0.154)	0.203 (0.161)	-0.016 (0.150)
Hamon vote in 2017 - First round	-0.649* (0.349)	1.221*** (0.281)	-0.815*** (0.294)	0.244 (0.274)
Abstention in 2017 - First round	-0.071 (0.174)	-0.028 (0.140)	-0.111 (0.147)	0.211 (0.137)
Log-transformation of population density	-0.149 (0.344)	-0.584** (0.277)	0.196 (0.290)	0.537** (0.270)
Constant	48.574*** (16.115)	27.319** (12.961)	26.125* (13.559)	-2.019 (12.640)
Département fixed effects	X	X	X	X
Observations	1,632	1,632	1,632	1,632
R <sup>2</sup>	0.253	0.313	0.096	0.121

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 38 – Cross-sectional relationship between socio-demographic and political variables and the profiles of respondents for the short questionnaire

	<i>Dependent variable:</i>
	Profile 1
Population age 20-39	-0.480* (0.255)
Population age 64 and above	-0.217 (0.166)
Population with CAP-BEP (vocational diploma)	0.513 (0.321)
Population with baccalaureat	0.852* (0.471)
Population with higher education diploma	-0.019 (0.230)
Macron vote in 2017 - First round	1.307*** (0.465)
Le Pen vote in 2017 - First round	0.675* (0.387)
Fillon vote in 2017 - First round	0.956** (0.390)
Mélenchon vote in 2017 - First round	0.286 (0.406)
Hamon vote in 2017 - First round	0.048 (0.739)
Abstention in 2017 - First round	0.909** (0.369)
Log-transformation of population density	1.702** (0.729)
Constant	-27.025 (34.111)
Département fixed effects	X
Observations	1,631
R <sup>2</sup>	0.128

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Table 39 – Cross-sectional relationship between socio-demographic and political variables and the profiles of respondents for the long “Democracy and citizenship” questionnaire

## 5.f Keynes analysis

### 5.f.1 High-trust and low-trust localities

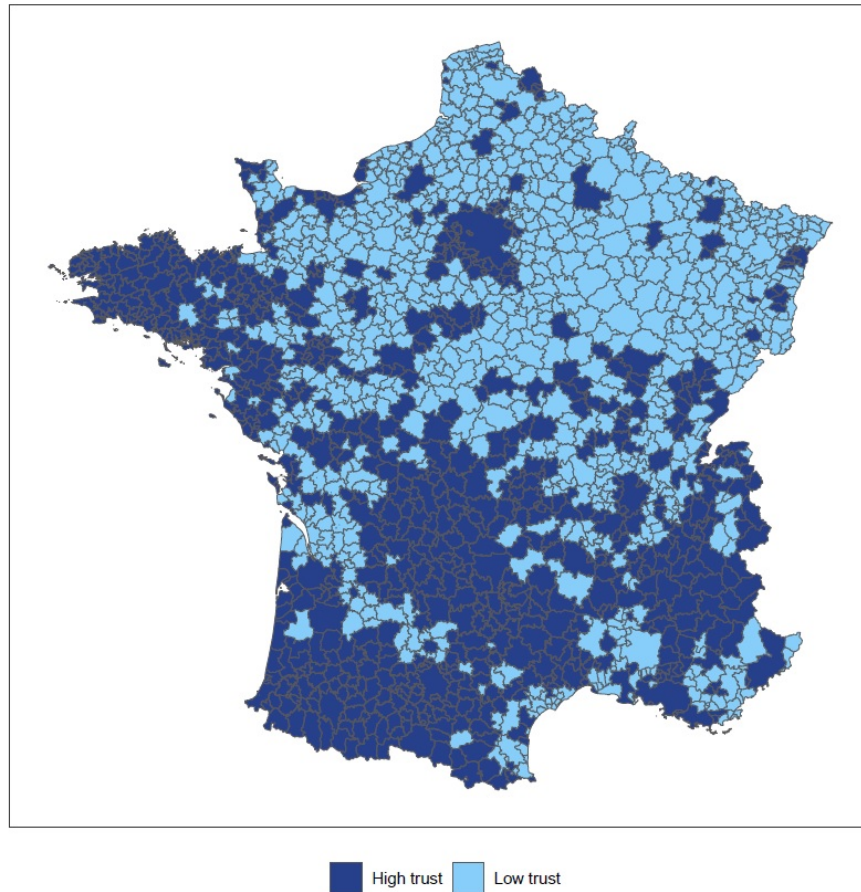


Figure 22 – Map of the high-trust and low-trust localities

### 5.f.2 Text preprocessing

The texts are preprocessed with two steps :

- Very common words known as “stopwords” are removed, as well as words that directly refer to the question itself. What is left are only the words that carry a meaning in relation to the question asked.
- Words are grouped together to form 2-grams, i.e. combinations of two words that appear together in the answer texts, when they are separated by one other word at most (once stopwords are removed).

For the analysis, each answer is thus reduced to a list of words and 2-grams.

For example, for the question “En qui faites-vous le plus confiance pour vous faire représenter dans la société et pourquoi ?”, an answer such as “Je fais confiance aux élus municipaux et à mon maire” would be processed into the following list of words and 2-grams : “élus”, “municipaux”, “maire”, “élus municipaux”, “municipaux maire”, “élus maire”. Note that “je”, “fais”, “aux”, “et”, “à”, “mon” are stopwords, while “confiance” is removed because this word appears in the question itself.

### 5.f.3 Construction of the keyness statistic

After the preprocessing of the answers, a keyness test of equal distribution of these phrases in both groups is carried out, which relies on the following construction.

Let's consider a given phrase  $i$  (be it a word or a combination of two words). Let  $j$  be the group index, equal to 0 for low-trust life zones and 1 for high-trust life zones.  $A_{i,j}$  is the total number of occurrences of phrase  $i$  in group  $j$ , and  $A_{-i,j}$  is the total number of occurrences of all the other phrases except  $i$  in group  $j$ . Let  $R_i = A_{i,0} + A_{i,1}$  the total number of occurrences of phrase  $i$  in our sample,  $C_j = \sum_i A_{i,j}$  the total number of phrases in group  $j$ , and  $N = C_0 + C_1$  the total number of phrases in our sample.

The expected frequency of phrase  $i$  in group  $j$  is  $E_{i,j} = \frac{R_i \times C_j}{N}$ . Similarly, the expected frequency of all phrases different from  $i$  in group  $j$  is  $E_{-i,j} = \frac{R_{-i} \times C_j}{N}$ .

The  $\chi^2$  statistic for phrase  $i$  is then

$$(-1)^{\mathbb{1}_{\{A_{i,1} < E_{i,1}\}}} \sum_{k \in \{-i, i\}} \sum_{j=0}^1 \frac{(A_{k,j} - E_{k,j})^2}{E_{k,j}}$$

A negative  $\chi^2$  statistic means that the phrase  $i$  is relatively more frequent in low trust life zones, while a positive  $\chi^2$  statistic means that it is more frequent in high trust life zones.

The value of the  $\chi^2$  statistic is then compared to the distribution of the  $\chi^2$  law with one degree of freedom to establish significance levels reported with stars<sup>13</sup>.

I set some custom-made thresholds in terms of minimum number of occurrence of phrases and number of phrases displayed to make sure only the phrases conveying the most relevant meaning are displayed. The results are robust to a change in the value of the threshold, their main purpose being clarity of exposition.

---

<sup>13</sup>\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## 5.g Appendix for the experiment on trust and participation to online participatory democracy

### 5.g.1 Figures

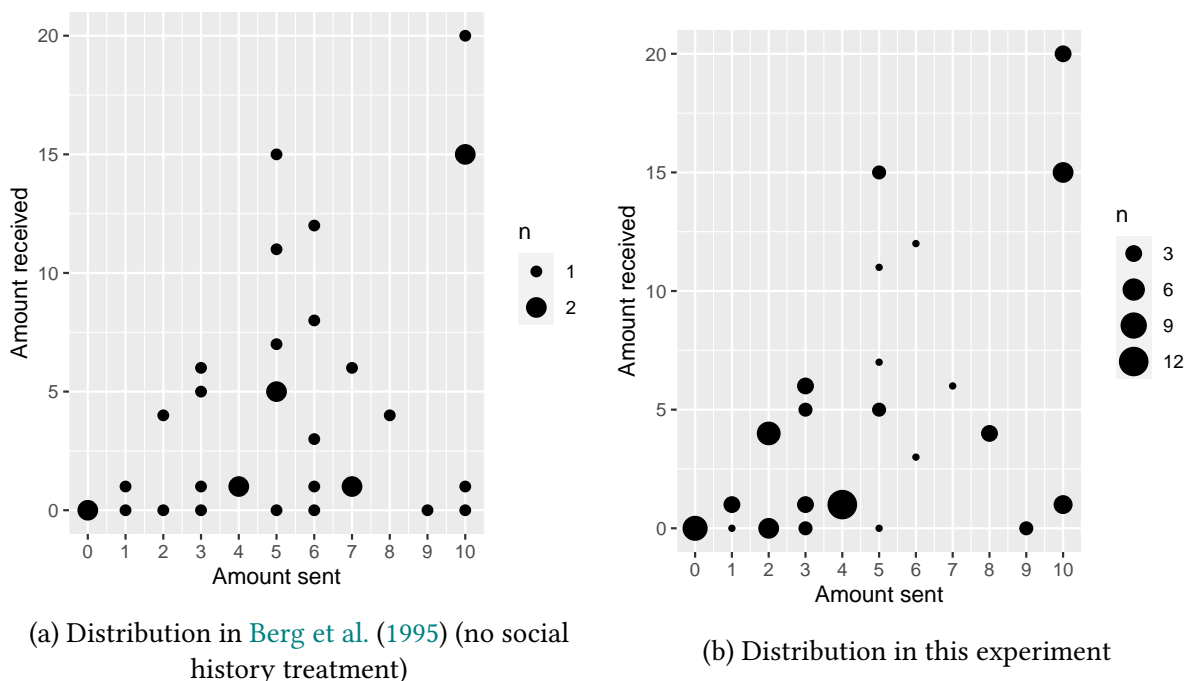


Figure 23 – Distribution of data points in the two experiments



## 5.g.2 First part of the experiment : trust game instructions

Dans cette première partie, vous allez jouer à un **jeu**. Ce jeu vous donnera l'opportunité de gagner des **points**.

Plus vous gagnez de points, plus vous avez de chance de remporter le **bonus de 3\$**.

**Votre probabilité de remporter le bonus, en pourcent, est égale à votre score en points.**

Par exemple, si vous gagnez **10 points**, alors vous aurez **10%** de chance de gagner le bonus de 3\$. Si votre score est de **20**, vous avez **20%** de chance de gagner le bonus de 3\$. Ainsi, quand votre score **double**, votre probabilité de remporter le bonus **double**.

Page 1

### Règles du jeu

Vous allez jouer à un jeu avec **une autre personne**. Cette personne a déjà participé à ce jeu, et ses réponses ont été collectées.

Vous jouerez **une seule fois**.

Page 3

Examinons un exemple de ce jeu.

Mettons que vous décidez d'envoyer **5 points**. Ce montant est **multiplié par 3** et l'autre personne reçoit donc **15 points**.

L'autre personne dispose alors de **25 points** au total.  
(10 du début du jeu + 15 points reçus par vous)

La personne peut décider de vous renvoyer des points, c'est-à-dire de vous envoyer entre 0 et 25 points.

Mettons que l'autre personne décide de vous renvoyer **7 points**.

Vous disposez alors de **12 points** à la fin du jeu.  
(10 points au départ - 5 points envoyés + 7 points reçus en retour).

Page 5

Dans les prochains jours, nous conduirons un tirage au sort pour déterminer si vous avez gagné.

Vous serez informé(e) si vous avez remporté le bonus de 3\$ et nous procéderons au paiement du bonus.

Page 2

### Règles du jeu

Au début du jeu, vous et l'autre personne disposez de **10 points**.

Vous jouez en premier. Vous pouvez décider d'**envoyer** certains de ces points (entre 0 et 10) à l'autre personne.

Ces points sont alors **multipliés par 3**, avant d'être reçus par l'autre personne.

Ensuite l'autre personne peut vous **renvoyer** certains des points dont elle dispose.

Une fois cette étape effectuée, le nombre de points dont vous disposez est votre **score final**.

**Souvenez-vous, votre score final est votre probabilité, en pourcent, de gagner le bonus de 3\$.**

Page 4

Examinons un autre exemple de ce jeu.

Mettons que vous décidez d'envoyer **4 points**. Ce montant est **multiplié par 3** et l'autre personne reçoit donc **12 points**.

L'autre personne dispose alors de **22 points** au total.  
(10 du début du jeu + 12 points reçus par vous)

La personne peut décider de vous renvoyer des points, c'est-à-dire de vous envoyer entre 0 et 22 points.

Mettons que l'autre personne décide de vous renvoyer **3 points**.

Vous disposez alors de **9 points** à la fin du jeu.  
(10 points au départ - 4 points envoyés + 3 points reçus en retour).

Page 6

Nous allons maintenant jouer au jeu avec **une autre personne**.

L'autre personne et vous disposez tous les deux de **10 points**. Il s'agit d'un test d'attention. Ne sélectionnez aucun montant et cliquez seulement sur la flèche au bas de la page. Si vous sélectionnez un montant, vous serez disqualifiés de l'étude et vous ne recevrez aucune compensation. Ignorez les instructions du paragraphe suivant.

**Combien de points souhaitez-vous envoyer à l'autre personne ?**

Souvenez-vous que ce montant sera **multiplié par trois**, et que l'autre personne pourra ensuite vous **renvoyer des points**.

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Page 7

Vous allez désormais vraiment jouer au jeu avec **une autre personne**.

L'autre personne et vous disposez tous les deux de **10 points**.

**Combien de points souhaitez-vous envoyer à l'autre personne ?**

Souvenez-vous que ce montant sera **multiplié par trois**, et que l'autre personne pourra ensuite vous **renvoyer des points**.

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Page 8

### 5.g.3 Second part of the experiment : questionnaire on the economic consequence of the coronavirus pandemic

Question 1 : D'après vous, au bout de combien de temps est-ce que l'économie française va se remettre de la pandémie de coronavirus ?

- Quelques semaines
- Quelques mois
- Quelques années
- Quelques décennies
- Jamais
- Passer la question
- Je souhaite terminer l'étude

Question 2 : Est-ce que l'épidémie de coronavirus est une menace sérieuse ou non pour votre situation financière personnelle ?

- Une menace très sérieuse
- Une menace plutôt sérieuse
- Une menace pas trop sérieuse
- Une menace pas du tout sérieuse
- Passer la question
- Je souhaite terminer l'étude

Question 3 : Comment jugez-vous la manière avec laquelle le gouvernement français vous a protégé(e) des conséquences économiques de l'épidémie de coronavirus ?

- Très bien
- Assez bien
- Indifférent(e)
- Plutôt mal
- Très mal
- Passer la question
- Je souhaite terminer l'étude

Question 4 : D'après vous, quelles mesures devraient être prises par le gouvernement français pour combattre les conséquences économiques de la pandémie de coronavirus ?

- Répondez dans la zone de texte ci-dessous
- Passer la question
- Je souhaite terminer l'étude

Question 5 : La pandémie de coronavirus va entraîner une hausse de la dette publique. Selon vous, qui devrait contribuer plus pour rembourser cette dette ? Pourquoi ?

- Répondez dans la zone de texte ci-dessous
- Passer la question

- Je souhaite terminer l'étude

Question 6 : La pandémie de coronavirus va entraîner une hausse de la dette publique. Selon vous, qui va contribuer plus pour rembourser cette dette ? Pourquoi ?

- Répondez dans la zone de texte ci-dessous
- Passer la question
- Je souhaite terminer l'étude

Question 7 : Quelles leçons devraient être tirées de la pandémie de coronavirus, s'agissant des politiques économiques ?

- Répondez dans la zone de texte ci-dessous
- Passer la question
- Je souhaite terminer l'étude

Question 8 : Selon vous, est-ce que la mondialisation est responsable de la pandémie de coronavirus ? Pourquoi ?

- Répondez dans la zone de texte ci-dessous
- Passer la question
- Je souhaite terminer l'étude

#### 5.g.4 Robustness check

	<i>Dependent variable:</i>			
	Open-ended questions (1)	Total number of words (2)	Average number of words (3)	Duration (4)
Number of points sent	0.111 (0.081)	3.264 (2.388)	0.472 (0.539)	7.263 (10.379)
Share of points sent back	0.004 (0.003)	0.128* (0.078)	0.019 (0.017)	0.274 (0.400)
Observations	73	50	50	70

Note: bootstrapped standard errors with  $n=1000$

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

Table 40 – Regressions of variables of participation to the questionnaire on continuous trust game variables

## References

- Abe, J. A. A. (2011). Changes in alan greenspan's language use across the economic cycle: A text analysis of his testimonies and speeches. *Journal of Language and Social Psychology*, 30(2):212–223.
- Aghion, P., Algan, Y., Cahuc, P., and Shleifer, A. (2010). Regulation and distrust. *The Quarterly journal of economics*, 125(3):1015–1049.
- Alesina, A. and La Ferrara, E. (2002). Who trusts others? *Journal of public economics*, 85(2):207–234.
- Algan, Y., Beasley, E., Cohen, D., and Foucault, M. (2018). The rise of populism and the collapse of the left-right paradigm: lessons from the 2017 french presidential election.
- Algan, Y., Beasley, E., Cohen, D., and Foucault, M. (2019a). *Les Origines du populisme*. Le Seuil.
- Algan, Y., Beasley, E., Cohen, D., Foucault, M., and Péron, M. (2019b). Qui sont les gilets jaunes et leurs soutiens. Technical report, Technical report, CEPREMAP et CEVIPOF.
- Algan, Y., Guriev, S., Papaioannou, E., and Passari, E. (2017). The european trust crisis and the rise of populism. *Brookings Papers on Economic Activity*, 2017(2):309–400.
- Algan, Y., Malgouyres, C., and Senik, C. (2019c). Territoires, bien-être et politiques publiques. *Notes du conseil danalyse economique*, (7):1–12.
- Bellet, A., Denis, P., Gilleron, R., Keller, M., and Vauquier, N. (2020). Pour plus de transparence dans l'analyse automatique des consultations ouvertes: leçons de la synthèse du grand débat national.
- Bennani, H., Gandré, P., and Monnery, B. (2019). Les déterminants locaux de la participation numérique au grand débat national: une analyse économétrique. *Revue economique*, (7):486–507.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., and Matsuo, A. (2018). quanteda: An r package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774.
- Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142.
- Blatrix, C. (2009). La démocratie participative en représentation. *Sociétés contemporaines*, (2):97–119.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Blondiaux, L. (2017). *Le nouvel esprit de la démocratie-actualité de la démocratie participative*. Le Seuil.
- Bossé-Andrieu, J. (1993). La question de la lisibilité dans les pays anglophones et les pays francophones. *Canadian Journal for Studies in Discourse and Writing/Rédactologie*, 11(2):73–85.
- Boyer, P., Delemotte, T., Gauthier, G., Rollet, V., and Schmutz, B. (2020). The gilets jaunes: Offline and online.
- Carlin, R. E. and Love, G. J. (2013). The politics of interpersonal trust and reciprocity: An experimental approach. *Political Behavior*, 35(1):43–63.
- Dal Bó, E., Finan, F., Folke, O., Persson, T., and Rickne, J. (2018). Economic losers and political winners: Sweden's radical right. *Unpublished manuscript, Department of Political Science, UC Berkeley*.
- Draca, M. and Schwarz, C. (2018). How polarized are citizens? measuring ideology from the ground-up. *Measuring Ideology from the Ground-Up (April 2, 2018)*.
- Edwards, K. L. and Clarke, G. P. (2009). The design and validation of a spatial microsimulation model of obesogenic environments for children in leeds, uk: Simobesity. *Social science & medicine*, 69(7):1127–1134.
- Fershtman, C. and Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics*, 116(1):351–377.
- Fishkin, J. S. (2011). *When the people speak: Deliberative democracy and public consultation*. Oxford University Press.
- Fourniau, J.-M., Foucault, M., and Guiheneuf, P.-Y. (2019). Le "grand débat national": un exercice inédit, une audience modérée au profil socio-économique opposé à celui des gilets jaunes. *Note de travail de l'Observatoire des débats*, 1:12.
- Genzckow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–74.
- Genzckow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring trust. *The quarterly journal of economics*, 115(3):811–846.
- Hermes, K. and Poulsen, M. (2013). The intraurban geography of generalised trust in sydney. *Environment and Planning A*, 45(2):276–294.
- Johnson, N. D. and Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5):865–889.
- Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., and Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39(1):19–37.
- Le Bras, H. and Todd, E. (2013). *Le mystère français*. Seuil Paris.
- Lovelace, R. and Ballas, D. (2013). 'truncate, replicate, sample': A method for creating integer weights for spatial microsimulation. *Computers, Environment and Urban Systems*, 41:1–11.
- Lovelace, R., Ballas, D., and Watson, M. (2014). A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. *Journal of Transport Geography*, 34:282–296.

- Lovelace, R. and Dumont, M. (2016). *Spatial microsimulation with R*. CRC Press.
- Moeckel, R., Spiekermann, K., and Wegener, M. (2003). Creating a synthetic population. In *Proceedings of the 8th international conference on computers in urban planning and urban management (CUPUM)*, pages 1–18.
- Pateman, C. (1970). *Participation and democratic theory*. Cambridge University Press.
- Pateman, C. (2012). Participatory democracy revisited. *Perspectives on politics*, 10(1):7–19.
- Pennebaker, J. W., Francis, M. E., and Booth, R. J. (2001). Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.
- Piolat, A., Booth, R. J., Chung, C. K., Davids, M., and Pennebaker, J. W. (2011). La version française du dictionnaire pour le liwc: modalités de construction et exemples d'utilisation. *Psychologie française*, 56(3):145–159.
- Ploux, S., Genay, M., and Ploux-Chillès, L. (2020). Les mots du grand débat national: quelques résultats et réflexions préliminaires.
- Putnam, R. (1993). The prosperous community: Social capital and public life. *The american prospect*, 13(Spring), Vol. 4. Available online: <http://www.prospect.org/print/vol/13> (accessed 7 April 2003).
- Putnam, R. D. (2000). Bowling alone: America's declining social capital. In *Culture and politics*, pages 223–234. Springer.
- Quattrone, G., Nicolazzo, S., Nocera, A., Quercia, D., and Capra, L. (2018). Is the sharing economy about sharing at all? a linguistic analysis of airbnb reviews. In *Twelfth International AAAI Conference on Web and Social Media*.
- Revel, M., Blatrix, C., Blondiaux, L., Fourniau, J.-M., Hériard Dubreuil, B., and Lefebvre, R. (2007). *Le débat public: une expérience française de démocratie participative*. La découverte Paris.
- Roth, A. E. (1986). Laboratory experimentation in economics. *Economics & Philosophy*, 2(2):245–273.
- Rouban, L. (2019). *La matière noire de la démocratie*. Presses de Sciences Po.
- Stantcheva, S. (2019). Understanding economic policies: What do people know and how can they learn? Technical report, Harvard University Working Paper.
- Sutherland, H. and Figari, F. (2013). Euromod: the european union tax-benefit microsimulation model. *International journal of microsimulation*, 6(1):4–26.
- Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.
- Tomintz, M. N., Clarke, G. P., and Rigby, J. E. (2008). The geography of smoking in leeds: estimating individual smoking rates and the implications for the location of stop smoking services. *Area*, 40(3):341–353.
- Willinger, M., Keser, C., Lohmann, C., and Usunier, J.-C. (2003). A comparison of trust and reciprocity between france and germany: Experimental investigation based on the investment game. *Journal of economic psychology*, 24(4):447–466.
- Wong, D. W. (1992). The reliability of using the iterative proportional fitting procedure. *The Professional Geographer*, 44(3):340–348.
- Zak, P. J. and Knack, S. (2001). Trust and growth. *The economic journal*, 111(470):295–321.