



HAL
open science

The impact on nudge acceptability judgments of framing and consultation of the targeted population

Ismaël Rafai, Arthur Ribailier, Dorian Jullien

► To cite this version:

Ismaël Rafai, Arthur Ribailier, Dorian Jullien. The impact on nudge acceptability judgments of framing and consultation of the targeted population. 2021. hal-03228638

HAL Id: hal-03228638

<https://hal.science/hal-03228638v1>

Preprint submitted on 18 May 2021

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.

The impact on nudge acceptability judgments of framing and consultation of the targeted population

Ismaël Rafai^{1,2,*}

Arthur Ribaillier²

Dorian Jullien^{3,2}

* *Corresponding author, ismael.rafai@gmail.com*

¹ CEE-M, Univ. Montpellier, CNRS, INRAE, Institut Agro, Montpellier, France.

² Université Côte d'Azur, CNRS, GREDEG

³ Centre d'Economie de la Sorbonne, Université Paris I - Panthéon Sorbonne

Abstract

The aim of this article is to better understand how judgments about nudge acceptability are formed and whether they can be manipulated. We conducted a randomized experiment to test whether acceptability judgments could be (i) more favourable when the decision to implement the nudges was made following a consultation with the targeted population and (ii) influenced by the joint framing of the nudge's goal and effectiveness (in terms of an increase in desirable behaviour vs. decrease in undesirable behaviour). We tested these hypotheses on various nudge scenarios and obtained mixed results that do not clearly support our hypotheses. A surprising result that calls for further work is that by mentioning that a nudge had been implemented through a consultation with the targeted population its acceptability could be lowered.

Keywords: behavioural public policies; nudges; acceptability; framing; consultation.

JEL: C91; D90 ; D04

Acknowledgment

This paper reports the results of an experiment conducted during Arthur Ribaillier's master thesis, entitled "L'acceptabilité des nudges digitaux", under Ismaël Rafai's supervision. The authors have not received any particular financial support and report no conflict of interest. The authors thank the LAPCOS (Université Côte d'Azur) for technical support, and Isabelle Milhabet, Daniel Priolo, Pierre Therouane, Guilhem Lecouteux, Agnès Féstré, and Mira Toumi for their valuable comments.

Introduction

Nudges are behavioural policy interventions that slightly modify the decisional context, without much affecting material incentives and without enforcing a particular choice, thus limiting the infringement of freedom and preserving autonomy. Nudges have been used all over the world in many domains such as health (e.g., displaying red or green coloured spots on food items to indicate healthiness), the environment (e.g., indicating your neighbours' consumption of electricity on your electricity bills), retirement (e.g., changing the default option from non-enrolment to enrolment in the pension plan provided by companies), among many others (see Thaler and Sunstein, 2008; Oliver, 2013; World Bank, 2015; OECD, 2017; Sanders et al., 2018a).

The ethical debate over nudges has been quite rich over the past fifteen years or so. A large portion of the debate has focused on whether or not nudges effectively preserve the liberty and the autonomy of the targeted population. Most contributions on this point have been philosophical and theoretical (see the review of Congiu and Moscati, 2020). However, there is a growing set of empirical studies that aim to contribute to this point by measuring the way people (i.e., not scholars with an interest in nudges) judge various nudges to be more or less acceptable (see the references cited below and in section 7.3 of Congiu and Moscati, 2020). Three main findings from these studies are that: (a) overall, most nudges are well accepted (e.g., people would support such policies or think that most nudges respect their liberty and autonomy), (b) there is a link between effectiveness and acceptability in the sense that the more people judge a nudge to be effective in changing behaviour, the more they tend to judge this nudge as being acceptable, and (c) acceptability judgments can be influenced by the framing of the evaluation tasks (e.g., a nudge tends to be judged as more acceptable when evaluated in isolation than when evaluated in comparison with alternative policies).

In this paper, we present an empirical study on the acceptability of nudges that has two related

aims. First, we aim to further our understanding of the connection between the last two points mentioned above by investigating whether the acceptability of a nudge is influenced by the framing of its effectiveness, i.e., whether presenting the effectiveness of a nudge in increasing desirable behaviour or in decreasing undesirable behaviour affects its acceptability. Second, we aim to better understand which aspects matter in the communicative structure at play in the elicitation of acceptability judgments. To do so we investigated the impact on acceptability judgments not only of the framing of the effectiveness of the nudge, but also of its goal (of increasing desirable behaviour vs. decreasing undesirable behaviour) and of whether the decision to implement the nudge is presented as having been made in consultation with representatives of the targeted population. This second aim is motivated by recent (and disparate) contributions to the ethical debate on nudges suggesting that their acceptability could be strengthened by taking more seriously the communicative structure between nudgers and their targeted population (Krijnen et al., 2017; McKenzie et al., 2018) or by involving more directly the latter in the decision process that leads to the implementation of nudges (John and Stoker, 2019).

Background

Our study is motivated by a set of contributions that have investigated the link between the effectiveness and the acceptability of nudges and that have highlighted the importance of taking seriously the communicative structure between nudgers and their targeted population.

Effectiveness and acceptability

It seems that a commonly held view in policy circles is (or at least was) that the acceptability of a behavioural intervention is a necessary condition for its effectiveness, i.e., interventions that are not judged as being acceptable by the public will not be effective (see, e.g., UK House of Lords Science and Technology Select Committee, 2011; Branson et al., 2015). Recent

empirical studies on the acceptability of nudges suggest that the link between effectiveness and acceptability is more complicated.

Firstly, the perceived effectiveness of a nudge is often found to be a very good predictor of its acceptability, even if that perception is inaccurate. For instance, Jung and Mellers (2016) found that nudges that imply little deliberation from the targeted population (such as switching default options) are perceived as being less acceptable and less effective than nudges that imply more deliberation (such as reminders), even though empirical studies have shown that the former tend to be more effective than the latter. This empirical link between perceived effectiveness and acceptability has been found in a number of studies (see, e.g., Petrescu et al., 2016; Bang et al., 2018; Cadario and Chandon 2018; Djupegot and Hansen, 2020; Gold et al., 2021).

Secondly, and maybe more crucially, when the effectiveness of a nudge is actually displayed explicitly during the evaluation task, the presence of this information tends to increase the acceptability of the nudge. For instance, Sunstein and Reisch (2019, chapter 7) asked people which of two nudges they preferred between one that implies little deliberation and another that implies more deliberation. Without information about effectiveness, they found the usual result that people prefer nudges that imply more deliberation. However, when they mentioned that the nudge that implies less deliberation was “significantly more effective” than the other, enough people switched their preference to make it the preferred choice of the majority. They also showed that when they mentioned specific numbers to illustrate the effectiveness of nudges, it had the same effect (but did not have a significantly larger effect). Similar results with different measures of acceptability (e.g., ratings on scales, willingness to pay) have been found in a number of other studies (see, e.g., Pechey et al., 2014; Arad and Rubinstein, 2018; Reynolds et al., 2019; Davidai and Shafir 2020; Rafiq, 2021).

Reynolds et al. (2020) provide a meta-analysis that shows that communicating effectiveness can increase public support for policies that aim at changing behaviour (whether or not these

policies are nudges). They nevertheless found mixed results which, according to them, call for further experimental studies in order to investigate more precisely which aspects of the communication structure have an effect on the judgments of the targeted population.

Communicative structure

Both Krijnen et al. (2017) and McKenzie et al. (2018) offer a broad set of arguments to defend the idea of involving implicit social interaction and communication between the nudger and the targeted population when implementing nudges. According to them, insufficient attention has been paid to the communicative structure of nudges, which includes the way the goal of the nudge is presented by the nudger and how that presentation affects the targeted population's understanding of it. Their main focus is on the effectiveness of nudges, i.e., how the framing of the nudger's goal impacts the effectiveness of the nudge. A number of studies have investigated how different goals behind a given nudge impact the acceptability of the nudge (Jung and Mellers, 2016; Steffel et al. 2016; Reisch and Sunstein 2016; Tannenbaum et al., 2017; Bang et al., 2018). However, to the best of our knowledge there are no studies on how, for a given nudge, different ways of framing the same goal can impact its acceptability.

Another, more explicit and *ex ante* form of communication between nudgers and their targeted population has been recently discussed by various scholars (John 2018; John and Stoker 2019; de Jonge et al. 2018; Sanders et al. 2018b). The general idea is to involve the targeted population directly in the decision-making process that underlies the implementation of a nudge (or other types of behavioural policy) through various forms of deliberative forums in which the exchange of ideas can take place. An implicit motivation behind these experiences of deliberative democracy seems to be that the nudges that will be subsequently implemented will be perceived as being more acceptable, which will in turn improve their effectiveness. While there is indeed evidence of the effectiveness of nudges that have been implemented as

such, we are not aware of evidence of their acceptability.

The present study

The empirical study that we detail below investigates the impact of two factors on people's acceptability judgments about nudges: (1) whether the decision to implement the nudge was made in consultation with representatives of the targeted population (mention of a consultation, no mention of a consultation) and (2) the framing of both the goal behind the nudge and its effectiveness (increase in desirable behaviour for both goal and effectiveness, decrease in undesirable behaviour for both goal and effectiveness). We tested two hypotheses:

Hypothesis 1: The mention of a consultation with representatives of the targeted population will increase the acceptability of a nudge.

Hypothesis 2: The framing of nudges' goals and effectiveness will impact their acceptability.

We expected that the mention of a consultation with representatives of the targeted population would increase the acceptability of a nudge because it decreases the arbitrariness that one can potentially perceive in a nudge that is imposed in a technocratic fashion. We also expected that the framing of both the goals and the effectiveness of a nudge would impact their acceptability, though we were not sure in which direction since there are mixed results in the literature and this type of *joint* framing had not been previously tested to the best of our knowledge.

In order to increase the generalizability of the potential effects, each subject had to evaluate the acceptability of four nudges, which all varied in several dimensions (identity of the nudger, identity of the targeted population, behavioural domain, and type of nudge). Three of the nudges were genuine policies that had already been implemented somewhere in the world and one was a fictitious but plausible one. Finally, we also collected individual data on gender, age, education and political opinions with no particular expectations of observing particular

interaction effects.

Experimental design

We conducted an online randomized experiment with 171 participants (81 male, 85 female, 5 NA)¹ aged from 17 to 84 ($M = 41.24$, $SD = 16.15$) during two weeks starting on 18 August 2020. We used a 2 (framing: increase in desirable behaviour *vs.* decrease in undesirable behaviour) x 2 (consultation: mention *vs.* no mention) between-subject design. The experiment was implemented using Qualtrics. Each participant first read an introduction explaining (1) that they were taking part in research conducted at our University, (2) what nudges are and that participants will evaluate the acceptability of four nudges, and (3) that responses are anonymous and personal information is not collected. Each participant was then asked to read the description of a nudge and rate its acceptability. Each participant followed this procedure for a total of four nudge scenarios. When one scenario had been rated, another scenario appeared, and it was not possible to go back to previous scenarios to change the acceptability rating. At the end of the experiment, participants answered questions about their gender, age, education and political opinions. Participants were not compensated for their participation in the study. Three of the four nudge scenarios were inspired by real nudges. The “coffee” scenario described a change in the default amount of sugar in drinks supplied by university coffee vending machines in order to reduce sugar consumption (inspired by Priolo et al., 2020). The “hotel” scenario described the communication to hotel clients of the share of previous clients who had chosen to reuse their towel in order to reduce water consumption by the hotel (inspired by Bohner and Schlüte, 2014). The “election” scenario described a government sending its citizens an encouraging SMS in order to reduce abstention in the election (inspired by Gerber and Rogers, 2009). The last and fictitious nudge scenario was the “company”

¹ Gender balance was not targeted by the experimenters.

scenario: it described a company that implemented a system for tracking work hours on their employees' computers to reduce the time spent surfing the internet. The four scenarios were presented to each participant in a random order. The scenarios were chosen to test our hypotheses on a large variety of nudges (see Table 1).²

Scenario	Domain	Source	Type of nudge	Target of the nudge
<i>Coffee</i>	Health (reduce sugar consumption)	University	Default option	Students, university staff, professors
<i>Hotel</i>	Environment and economic (reduce water and energy consumption)	Hotel	Social comparison	Clients
<i>Election</i>	Democracy (reduce abstention)	Government	Positive message	Citizens
<i>Company</i>	Economic (increase productivity)	Consulting company	Salient information	Employees

Table 1: Characteristics of the four nudge scenarios.

These scenarios present diversity (i) in the type of nudge evaluated (default option, social comparison, positive message, salient information), (ii) in the source of the nudge (university, company, government, employer), (iii) in the target of the nudge (people working or studying at the university, consumers, citizens, employees), (iv) in the behavioural domain of the nudge (health, environment, democracy, economic). Variability in the scenarios was not introduced to directly test which nudge characteristics are potential moderators of the framing and consultation effects, but rather to stress the robustness of potential effects in numerous

² See the appendix 1 for the full scenarios.

situations. Our design was not adapted to attribute a difference in acceptance and effect heterogeneity to any particular nudge characteristic (except for the framing of the nudges and the presence of consultation, which were manipulated), since these characteristics always vary in more than one dimension from one scenario to another, preventing rigorous analysis due to confounding variables. All scenarios were constructed with a similar structure:

1. Introduction of the behavioural domain: We describe the domain in which a behavioural intervention is justified.
2. Goal [*frame ++ / frame --*]: We describe the goal of the nudger in terms of an increase in desirable behaviour (*frame ++*) or in terms of a decrease in undesirable behaviour (*frame --*).
3. Decision to implement the nudge [*mention of a consultation, no mention of a consultation*]: We either mention that the implementation of the nudge was decided in consultation with representatives of the targeted population (*mention of a consultation*), or we do not mention anything about how that decision was taken (*no mention of a consultation*).
4. Description of the nudge: We describe the type of nudge and explain why this nudge could be effective.
5. Effectiveness of the nudge [*frame ++ / frame --*]: We communicate the effectiveness of the nudge, in terms of an increase in desirable behaviour (*frame ++*) or in terms of a decrease in undesirable behaviour (*frame --*).

For each participant, we therefore manipulate two dimensions of the scenarios. The first dimension is the frame we used to describe the nudge scenario. Both the goal of the nudge and its effectiveness are described either in terms of an increase in desirable behaviour (*frame ++*) or in terms of a decrease in undesirable behaviour (*frame --*). We decided to manipulate the goal and the effectiveness of the nudge simultaneously to always present them in the same

frame (mixing the framing of the goal and effectiveness would have been very confusing). The other dimension is whether the targeted population had a voice in the decision process that led to the implementation of the nudge (*mention of a consultation*) or not (*no mention of a consultation*). Table 2 summarizes the relevant differences between treatments. To avoid spillover effects, we decided to present all the scenarios to each participant in the same condition (i.e. same combination of frame and consultation).

	Scenarios:	Coffee	Hotel	Election	Company
Goal	frame + +	To increase the number of users who do not take sugar in their drinks	To increase the number of customers reusing their towels	To increase the participation rate	To spend more time in the day on work-related activities
	frame - -	To decrease the number of users who take sugar in their drinks	To decrease the number of customers who do not reuse their towels	To decrease the abstention rate	To spend less time in the day on non-work-related activities
Consultation	no mention of consultation	The president of the university has decided	The hotel management has decided	The Minister of Interior has decided	The manager of the company has decided
	mention of a consultation	The president of the university has decided, following the proposal of the users representative council	After discussion with former clients , the hotel management has decided	The Minister of Interior, after consultation with citizen representatives , has decided	The manager of the company, on the advice of the works council , has decided
Effectiveness	frame + +	80% of users do not take sugar in their drink, that is 60% more than before	75% of users reuse their towel several days in a row during their stay, this is 25% more than before	60% participation , this is 10% more than the forecast	Work time dedicated to work-related activities increased by 30 minutes per day
	frame - -	20% of users take sugar in their drink, that is 60% less than before	25% of users do not reuse their towel several days in a row during their stay, this is 25% less than before	40% abstention , this is 10% less than the forecast	Work time dedicated to non-work-related activities decreased by 30 minutes per day.

Table 2: Experimental manipulation in each nudge scenario

Our dependent variable was obtained from the acceptability scale proposed by Tannenbaum et al. (2017), and translated into French by Priolo et al. (2020) (see appendix 2). Tannenbaum

(2017)'s acceptability scale is composed of the following items:

1. How much do you support this approach to policy?
2. How much do you oppose this approach to policy? (R)
3. How ethical is this approach to policy?
4. How manipulative is this approach to policy? (R)
5. How unethical is this approach to policy? (R)
6. How coercive is this approach to policy? (R)

Each item was rated on a 5-point Likert scale (1=Not agreeing at all, 5=Totally agreeing). “(R)” indicates reversed items. For each scenario, we summed up the scores of all items (for the reversed items we added 6 and subtracted their score) to compute a single *Acceptability index*.

We used Acceptability indexes (one for each scenario) as the dependent variables of linear regression models with framing and consultation treatments as independent variables. We estimated these models with and without including the individuals' characteristics as controls.

As control variables, we collected information at the end of the experiment about gender (masculine, feminine, other), age and political opinion (from 1 = far left to 5 = far right). At the time of the experiment, we did not expect particular interactions between these controls and the effect of framing and consultation but aimed to investigate potential interaction as an exploratory analysis. In total, 158 out of 171 participants provided complete answers to the control variables.³

Results

³ While in theory, Null Hypothesis Testing results are mathematically valid only for confirmatory analyses, exploratory analyses are useful to identify future hypotheses to be tested. See Jaeger and Halliday (1998) for more discussion on the distinction between exploratory and confirmatory analyses.

Data and the R-script statistics are publicly available on osf.⁴

Internal validity and correlation between acceptability indexes

We first tested the internal validity of the acceptability scale across the fourth scenario. We found a Cronbach's alpha ranging from 0.741 to 0.893, suggesting that items in the acceptability scale capture a similar concept. Table 3 summarizes mean and standard deviation of the acceptability index and acceptability scale Cronbach alpha. Figure 1 shows the distribution of the acceptability index. Note that the maximum acceptability rating for a nudge is 30 (6 items, each rated on a maximum of 5 points).

Scenarios	Coffee	Hotel	Election	Company
Mean	25.18	23.60	20.67	17.90
Standard deviation	4.36	5.04	5.82	6.70
Cronbach's alpha	0.751	0.791	0.876	0.893

Table 3: Interval validity of the acceptability scale and mean and standard deviation of the Acceptability index in the four nudge scenarios.

Note: Differences between mean acceptability indexes across scenarios are all significant at the 0.001 level (two-tailed paired t.test).

⁴ <https://osf.io/69uac/>

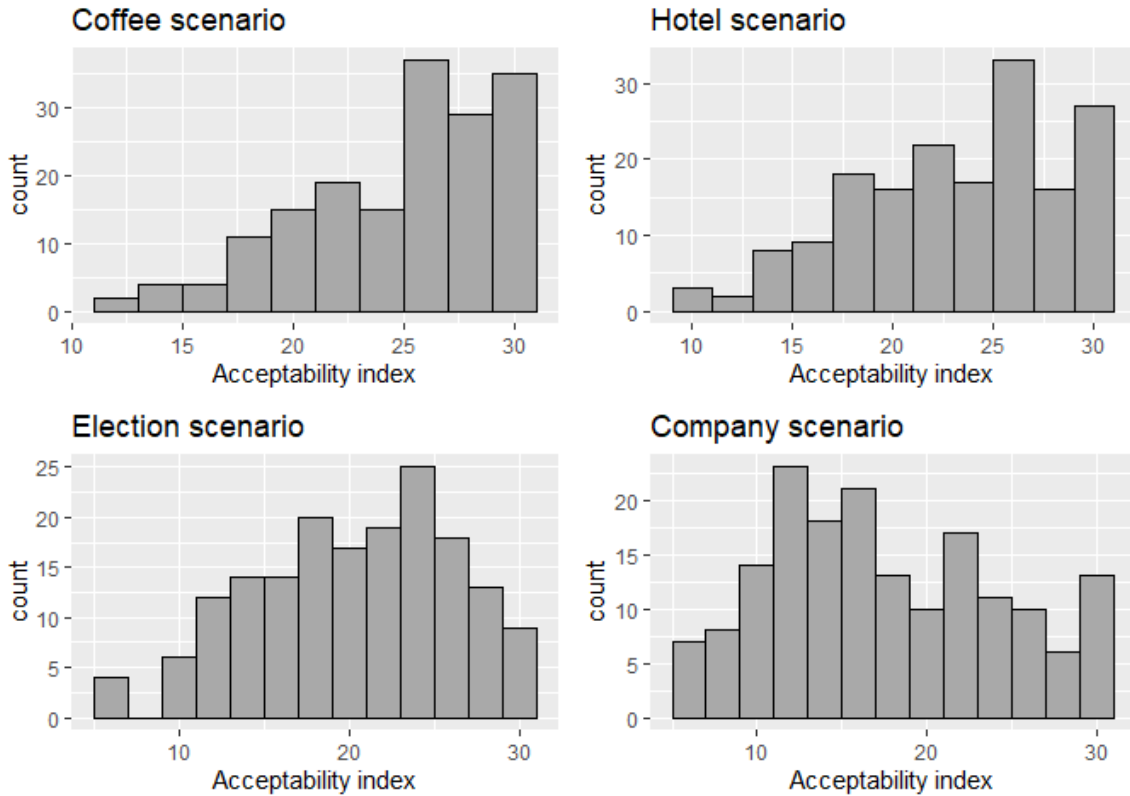


Figure 1: Distribution of the Acceptability index.

Note: Distribution of the acceptability index for the coffee (top left), hotel (top right), election (bottom left) and company (bottom right) scenarios.

As we proposed high variability across scenarios, we investigated the correlation between the acceptability index of each scenario (see Table 4). The Pearson correlation coefficients between the acceptability indexes are all positive (r between 0.216 and 0.362) and statistically significant (at least at the 1% level), except the acceptability index of the company scenario which is significantly correlated only with the hotel scenario ($r = 0.362, p < 0.001$) and correlated at the 10% level with the coffee scenario ($r = 0.129, p = 0.092$).

	Election	Company	Hotel
Coffee	0.216 **	0.129 ·	0.358 ***
Hotel	0.242 **	0.362 ***	
Company	0.057		

Table 4: Correlation between the acceptability of the different scenarios.

Note: Pearson's correlation test. ***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$; · $p < 0.1$

Confirmatory results

We conducted separate OLS regressions with each scenario's acceptability measures as dependent variables, and with "frame + +" and consultation as independent variables. We ran regressions with and without control variables (including gender, age, political opinion, level of education, and the order of the scenarios within the experiment). The results of the regressions are summarized in Table 5 (models 1, 3, 5, 7).

We found negative significant effects for the positive frame and for consultation in the coffee and in the election scenarios (in the coffee scenario (model 1): *Cohen's d* = -0.181, $p = 0.0190$ for the positive frame and *Cohen's d* = -0.208, $p = 0.0072$ for consultation; in the election scenario (model 3): *Cohen's d* = -0.158, $p = 0.0410$ for the positive frame and *Cohen's d* = -0.182, $p = 0.0183$ for consultation). Interactions between the two treatments are positive and comparable to the direct effects, but significant only at the 10% level for the coffee scenario (*Cohen's d* = 0.139, $p = 0.0710$) and at the 5% level for the election scenario (*Cohen's d* = 0.174, $p = 0.0239$). To summarize, we found negative effects on nudge acceptability in the election and coffee scenarios for both positive frame and consultation, but those effects do not sum up. We found no significant effects for the hotel and company scenarios.

	Scenario's acceptability							
	Coffee		Hotel		Election		Company	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Frame + +	-2.130*	-2.572*	-0.492	-1.596	-2.498*	-2.263	-0.649	-2.579
	(0.899)	(0.999)	(1.067)	(1.156)	(1.212)	(1.332)	(1.413)	(1.408)
Consultation	-2.466**	-3.196**	-1.343	-1.791	-2.912*	-2.757*	0.536	-0.919
	(0.907)	(1.010)	(1.075)	(1.155)	(1.222)	(1.347)	(1.424)	(1.419)
Consult. × Frame +	2.399	3.149*	1.743	2.514	4.057*	3.177	1.939	4.882*
	(1.320)	(1.447)	(1.566)	(1.677)	(1.780)	(1.948)	(2.074)	(2.050)
Men		-1.519*		-0.254		0.663		-0.520
		(0.707)		(0.810)		(0.950)		(0.982)
Age		-0.009		-0.035		-0.080*		0.070*
		(0.025)		(0.028)		(0.033)		(0.035)
(right) political side		-0.106		1.434**		0.690		2.603***
		(0.460)		(0.524)		(0.604)		(0.634)
Constant	26.709***	30.512***	24.018***	24.334***	22.182***	25.029***	17.491***	9.603**
	(0.575)	(2.273)	(0.682)	(2.656)	(0.775)	(2.947)	(0.903)	(3.087)
Observations	171	158	171	158	171	158	171	158
R ²	0.060	0.108	0.011	0.095	0.041	0.106	0.018	0.269
Adjusted R ²	0.043	0.028	-0.007	0.013	0.024	0.025	0.00001	0.203
Residual Std. Error	4.264	4.309	5.057	4.912	5.748	5.749	6.698	5.978
F Statistic	3.540*	1.342	0.597	1.160	2.383	1.316	1.000	4.076***

Table 5: OLS results. Acceptability across the scenarios

Note: standard errors in parentheses. · $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Control variables indicating the order of the scenario within the experiment and the participant's level of education are not significant at the 10% level in any regression and are masked for the sake of clarity. Difference in the number of observations is explained by participants who did not answer the demographic questions.

Introducing control variables allows us to stress the robustness of these results. In the coffee scenario, the significance of framing (*Cohen's d* = - 0.205, $p = 0.0111$) and consultation (*Cohen's d* = - 0.252, $p = 0.0019$) are unchanged and the interaction terms became significant at the 5% level (*Cohen's d* = 0.173, $p = 0.0312$). In the election scenario, the effect of a consultation remains significant at the 5% level (*Cohen's d* = - 0.163, $p = 0.0425$) and the effect of framing is now significant only at the 10% level (*Cohen's d* = - 0.135, $p = 0.1381$), but the interaction effect is not more statistically significant ($p = 0.1050$). For the company scenario, we found a significant positive interaction effect (*Cohen's d* = 0.190, $p = 0.0185$) and

a negative effect of the positive frame, significant at the 10% level (*Cohen's d* = - 0.146, $p=0.0690$). Overall, these results can be considered as mixed and not as clearly supporting our hypotheses.

Explanatory results

Besides the reported confirmatory results, this experiment allows us to investigate to what extent the individuals' characteristics are predictors of the acceptability of the nudges. At the time of the study, we did not formulate precise hypotheses on this point and we thus present these results as exploratory. We observed that in the coffee scenario, participants who identified as men found the nudge less acceptable than those who identified as women (*Cohen's d* = - 0.171, $p = 0.0334$). Individuals with more right-wing political opinions find the hotel and company scenarios more acceptable (*Cohen's d* = 0.218, $p = 0.007$ for the hotel scenario and *Cohen's d* = 0.326, $p < 0.001$ for the company scenario). Older participants judge the election scenario less acceptable (*Cohen's d* = - 0.197, $p = 0.0145$), but judge the company scenario more acceptable than their younger counterparts (*Cohen's d* = 0.161, $p = 0.0445$). We found no significant effect of the level of education and of the order in which the scenarios were presented.

Discussion

We found that, on average, all of the nudges were judged as being acceptable as acceptability ratings for each nudge were above half of the maximum possible acceptability rating, i.e., were above 15 for a maximum possible rating of 30. This is consistent with the general tendency observed in acceptability studies, in which most nudges are judged to be well accepted (especially when information about their effectiveness is explicit, as already discussed above).

We did not find general support for our hypotheses that the acceptability of a nudge would be influenced by the joint framing of its goal and effectiveness, and also be positively influenced

by mentioning that it was implemented following consultation with representatives of the targeted population. More precisely, we found no support for both hypotheses in two nudge scenarios: the nudge through social comparison in the hotel scenario and the nudge through salient information in the company scenario. We offer the following speculative explanation for such a lack of effects. Notice that these two scenarios are the only two that involve a private organization. One could argue that because there is a clear goal behind most decisions implemented by private organizations, namely a profit motive, people are less inclined to infer tacit information from the way information is communicated, i.e., from the framing of the goal and effectiveness of the nudge, so that such framing does not impact their acceptability judgments. As for the lack of effect on acceptability of the mention of a consultation, this could be explained, in line with the ideals of deliberative democratic theory, by the fact that it is participation itself in the consultation (i.e., not its mere mention) that is a transformative experience which then impacts individuals' judgments (see Rosenberg, 2007).

We found mixed support for our two hypotheses in the remaining two nudge scenarios (the nudge through a change in the default option in the coffee scenario and the nudge through an SMS reminder in the election scenario). We found in these two cases that presenting both the goal and the effectiveness of the nudge in decreasing undesirable behaviour (in sugar consumption or in abstention) had a positive impact on acceptability ratings (significantly so in the coffee scenario and weakly significantly so in the election scenario). These results are in line with the seminal results of Meyerowitz and Chaiken's (1987) that health prevention campaigns that highlight the bad consequences of inaction tend to be more effective than the ones that highlight the good consequences of undertaking health-improving action. This tendency is traditionally explained by loss aversion (information about undesirable consequences are represented as losses and therefore have more impact than information about desirable consequences which are represented as gains; see Drouin et al., 2018, p.215 for a

concise discussion).

We also found (still in these same two cases) that mentioning the presence of a consultation with the targeted population in the decision process to implement the nudge had a negative impact on acceptability ratings (again, significantly so in the coffee scenario and weakly significantly so in the election scenario). This is a rather surprising result that plainly contradicts our first hypothesis (that mentioning a consultation would increase acceptability). We can speculate, based on insights from contributions on deliberative democratic processes, that people who initially judge these nudges to not be very acceptable polarize their judgment if they learn that the nudges were implemented through a consultation. People who do not necessarily care for these nudges might be annoyed by the deployment of what they might consider to be “meaningless consultation” (John, 2018, p.125) or processes that “place a high burden on citizens in terms of their time” (ibid, p.126) for relatively unimportant stakes. Or it can be that the mention of the consultation led people to not only judge the acceptability of the nudges but also devalue the contradictory judgments of those who participated to the consultation (Rosenberg, 2007, p.343). In any case, this negative effect for the acceptability of the mention of a consultation with representatives of the targeted population clearly calls for further studies.

We conclude by discussing our exploratory results. We found that women judged the coffee scenario more acceptable than men (we did not observe gender differences for the other scenarios). This is in line with standard results supporting a higher concern for health in women compared to men (Aliaga, 2002; Bertakis et al., 2000). We also found that older participants judged the election scenario to be less acceptable than their younger counterparts. We can cynically speculate that older people might be more disillusioned by democratic participation than younger people (who have had less opportunity to be disappointed by politicians). Finally, we found that participants with more right-wing political opinions judge the company scenario

to be more acceptable. Previous studies have shown that the political opinions of people positively influence their acceptability judgments about nudges when the nudger and/or the political valence of the nudge is congruent with these political opinions. For instance, the Bush Administration or nudges that simplify the procedures to obtain tax breaks for high-income individuals are congruent with conservative political opinions (for a general discussion of these results see Sunstein and Reisch, 2019, chapter 3). Arguably, of our four nudges, the one in the company scenario has both the nudger (the manager of the company) and the political valence of the nudge (increased productivity at work) that are the most congruent with more right-wing political opinions.

References

- Aliaga, C. (2002), Women more attentive to their health than men, Household Living Conditions division, *Insee*, N° 869.
- Arad, A., & Rubinstein, A. (2018). The people's perspective on libertarian-paternalistic policies. *The Journal of Law and Economics*, 61(2), 311-333.
- Bang, H. M., Shu, S. B., & Weber, E. U. (2018). The role of perceived effectiveness on the acceptability of choice architecture. *Behavioural Public Policy*, 1-21. doi:10.1017/bpp.2018.1
- Bertakis, K. D., Azari, R., Helms, L. J., Callahan, E. J., & Robbins, J. A. (2000). Gender differences in the utilization of health care services. *Journal of family practice*, 49(2), 147-147.
- Bohner, G., & Schlüter, L.E. (2014) A Room with a Viewpoint Revisited: Descriptive Norms and Hotel Guests' Towel Reuse Behavior. *PLoS ONE* 9(8): e104086.
- Branson, C., Duffy, B., Perry, C., & Wellings, D. (2012). Acceptable behaviour: Public opinion on behaviour change policy. *Ipsos MORI, London*.
- Cadario, R., & Chandon, P. (2018). Which healthy eating nudges work best? A meta-analysis of behavioral interventions in field experiments. *Appetite*, 130, 300-301.
- Congiu, L., & Moscati, I. (2021). A review of Nudges: Definitions, Justifications, Effectiveness. Available at SSRN: <https://ssrn.com/abstract=3728512>
- Davidai, S., & Shafir, E. (2020). Are 'nudges' getting a fair shot? Joint versus separate evaluation. *Behavioural Public Policy*, 4(3), 273-291.
- De Jonge, P., Zeelenberg, M., & Verlegh, P. W. (2018). Putting the public back in behavioral public policy. *Behavioural Public Policy*, 2(2), 218-226.
- Djupegot, I. L., & Hansen, H. (2020). If it works, I like it: consumer acceptance of food-related nudging. *Journal of International Food & Agribusiness Marketing*, 32(4), 361-378.
- Drouin, O., Young, M., & King, N. (2018). Message framing and counseling of parents on children's physical activity—an experimental study. *Health Psychology and Behavioral Medicine*, 6(1), 214-225.
- Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: public attitudes towards overt and covert nudges. *Judgment & Decision Making*, 8(3), 202-213.
- Gerber, A. S., & Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1), 178-191.
- Gold, N., Lin, Y., Ashcroft, R., & Osman, M. (2021). 'Better off, as judged by themselves': Do people support nudges as a method to change their own behavior? *Behavioural Public Policy*, doi:10.1017/bpp.2020.6.

- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, 6(3), 439-453.
- Jaeger, R. G., & Halliday, T. R. (1998). On confirmatory versus exploratory research. *Herpetologica*, S64-S66.
- John, P. (2018). *How Far to Nudge?*. Edward Elgar Publishing.
- John, P., & Stoker, G. (2019). Rethinking the role of experts and expertise in behavioural public policy. *Policy & Politics*, 47(2), 209-226.
- Jung, J. Y., & Mellers, B. A. (2016). American attitudes toward nudges. *Judgment & Decision Making*, 11(1), 62-74.
- Krijnen, J. M. T., Tannenbaum, D., & Fox, C. R. (2017). Choice architecture 2.0: Behavioral policy as an implicit social interaction. *Behavioral Science & Policy*, 3(2), 1–18.
- McKenzie, C. R., Sher, S., Leong, L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics*, 5(3-4), 337-360.
- Meyerowitz, B. E., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of personality and social psychology*, 52(3), 500.
- OECD (2017). *Behavioral Insights and Public Policy: Lessons from Around the World*. OECD Publishing, Paris.
- Oliver, A. (Ed.). (2013), *Behavioural public policy*, Cambridge University Press.
- Osman, M., Fenton, N., Pilditch, T., Lagnado, D., & Neil, M. (2018). Whom Do We Trust on Social Policy Interventions?. *Basic and Applied Social Psychology*, 40(5), 249-268.
- Pechey, R., Burge, P., Mentzakis, E., Suhrcke, M., & Marteau, T. M. (2014). Public acceptability of population-level interventions to reduce alcohol consumption: a discrete choice experiment. *Social science & medicine*, 113, 104-109.
- Petrescu, D. C., Hollands, G. J., Couturier, D. L., Ng, Y. L., & Marteau, T. M. (2016). Public acceptability in the UK and USA of nudging to reduce obesity: the example of reducing sugar-sweetened beverages consumption. *PLoS One*, 11(6).
- Priolo, D., Milhabet, I., Bertolino, M., Juille, T., Jullien, D., Lecouteux, G., Rafäi, I., and Thérouanne, P. (2020). Would you accept some coffee with your sugar? a natural field experiment on the efficiency and acceptability of setting zero sugar by default in coffee vending machines. Accepted in Principle in *Comprehensive Result in Social Psychology*.
- Rafiq, S. (2021). How much is a calorie worth? A study of willingness to pay for calorie labels in restaurant menus. *Behavioural Public Policy* doi:10.1017/bpp.2019.8.
- Reisch, L. A., & Sunstein, C. R. (2016). Do Europeans like nudges? *Judgment and Decision*

making, 11(4), 310-325.

Reynolds, J. P., Archer, S., Pilling, M. A., Kenny, M., Hollands, G. J., & Marteau, T. M. (2019). Public acceptability of nudging and taxing to reduce consumption of alcohol, tobacco, and food: A population-based survey experiment. *Social Science & Medicine*, 236, 112395.

Reynolds, J. P., Stautz, K., Pilling, M., van der Linden, S., & Marteau, T. M. (2020). Communicating the effectiveness and ineffectiveness of government policies and their impact on public support: A systematic review with meta-analysis. *Royal Society open science*, 7(1), 190522.

Rosenberg, S. W. (2007). Rethinking democratic deliberation: The limits and potential of citizen participation. *Polity*, 39(3), 335-360.

Sanders, M., Snijders, V. & Hallworth, M. (2018a). Behavioural science and policy: Where are we now and where are we going? *Behavioural Public Policy*, 2(2), 144-167.

Sanders, M., Snijders, V. & Hallworth, M. (2018b). Response to responses. *Behavioural Public Policy*. 2(2). 263-269.

Steffel, M., Williams, E. F., & Pogacar, R. (2016). Ethically deployed defaults: Transparency and consumer protection through disclosure and preference articulation. *Journal of Marketing Research*, 53(5), 865-880.

Sunstein, C. & Reisch, L. (2019). *Trusting Nudges. Toward a Bill of Right for Nudging*. Routledge

Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, 31(3), 355-366.

Tannenbaum, D., Fox, C. R., & Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behaviour*, 1(7), 0130, 1-7.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

UK House of Lords, Science and Technology Select Committee (2011), *Behaviour change (Second report)*. London, United Kingdom.

World Bank (2015), *World Development Report 2015: Mind, Society, and Behavior*, Washington DC: World Bank.

Appendices

Appendix 1: Nudge scenarios

Scenario 1: default option in coffee-vending machine

Translation (by the authors):

“According to WHO, sugar consumption is excessive in France. In a French University, it has been observed that 80% of the users of a coffee-vending machine put sugar in their drinks and 20% of users do not put sugar in their drinks. To **[frame ++: increase the number of users taking their drink without sugar / frame - -: decrease the number of users taking their drink with sugar]**, the president of the University has decided **[consultation=1: following the proposal of the users representative council]** to implement a default option prevention policy that does not limit the freedom of users. It has been decided to change the default sugar quantity in the University coffee-vending machines. The machines were programmed to offer a quantity of “3 out of 5” sugar doses and have now been programmed to offer a quantity of “0 out of 5”. The user remains free to change the sugar quantity. It has been proven that changing the default option has an impact on behaviour. Consequently, because of this policy, **[frame ++: 80% of users do not take any sugar in their drink, that is 60% more than before / frame - -: 20% of users take sugar in their drink, which is 60% less than before].”**

Original French version (experimental material):

D'après l'OMS, la consommation de sucre en France est trop élevée. Notamment, dans les distributeurs de boisson chaude d'une Université française, il a été observé que 80% des usagers (c'est à dire à la fois les étudiants et le personnel) prenaient du sucre dans leurs boissons et que 20% prenaient leurs boissons sans sucre.

Afin d'augmenter le nombre d'usagers qui prennent leurs boissons sans sucre, le président de l'Université a décidé de mettre en place une politique de prévention qui ne nuise pas à la liberté des usagers.

Ainsi il a été décidé de changer la quantité de sucre proposée par défaut dans les distributeurs de boissons chaudes de l'Université. Les machines, qui étaient précédemment programmées de façon à proposer par défaut "3 sucres sur 5" ont été programmées pour proposer par défaut "0 sucre sur 5", l'utilisateur gardant toujours la possibilité de changer la quantité de sucre lors de la commande.

En effet, il a été prouvé qu'un changement d'option par défaut permettait d'agir efficacement sur les comportements.

Ainsi, grâce à la mise en place de cette politique, 80% des usagers prennent désormais leurs boissons sans sucre, soit 60% en plus que précédemment.

Figure A1: Screenshot of the Coffee scenario (frame ++ and no consultation treatments).

Scenario 2: hotel

“The daily replacement of towels in hotels is an added comfort for many customers. However, this service is not considered essential by most of them and represents a significant cost for hoteliers and for the environment (expenditure of water and energy). A hotel on the French Riviera therefore prefers to ask customers, when they arrive, whether or not they agree to reuse their towels for several days in a row. It has been observed that 50% of hotel guests choose to reuse the same towel several days in a row. In order to **[fame + +: increase the number of clients reusing their towels / fame - -: decrease the number of clients not reusing their towels]** the management has decided to introduce **[consultation=1: after discussion with former clients]** a policy of persuasion that would not interfere with the freedom of clients. Thus, at the time of booking and the day before the start of their stay, customers receive the following message by email; “In this hotel, one in two customers have chosen to protect the environment by reusing their towels for several days in a row.” This was decided because it has been proven that social comparison is an effective way to change behaviour. Thanks to the implementation of this policy, **[fame + +: 75% of customers reuse their towel for several days in a row during their stay, this is 25% more than before / fame - -: 25% of customers do not reuse their towel for several days in a row during their stay, this is 25% less than before.]”**

Le remplacement quotidien des serviettes dans les hôtels apporte un confort important pour de nombreux clients. Cependant, ce service n'est pas jugé comme indispensable par tous et représente un coût non-négligeable pour les hôteliers et pour l'environnement (dépense d'eau et d'énergie). Un hôtel de la Côte d'Azur préfère donc demander, lors de l'arrivée des clients, s'ils souhaitent ou non réutiliser leurs serviettes plusieurs jours de suite. Il a été observé que 50% des clients de l'hôtel font le choix de réutiliser la même serviette plusieurs jours de suite.

Afin d'augmenter le nombre de clients réutilisant leurs serviettes, la direction a décidé de mettre en place une politique de persuasion qui ne nuirait pas à la liberté des clients.

Ainsi, au moment de la réservation et la veille du début du séjour, les clients reçoivent le message suivant par courriel :

“Dans cet hôtel, un client sur deux a fait le choix de protéger l'environnement en réutilisant sa serviette plusieurs jours de suite”.

En effet, il a été prouvé qu'un message comparant les individus avec ceux abordant le comportement positif était un moyen efficace pour inciter les autres à adopter le même comportement.

Ainsi, grâce à la mise en place de cette politique, 75% des clients réutilisent leur serviette plusieurs jours de suite durant leur séjour, soit 25% en plus que précédemment.

Figure A2: Screenshot of the hotel scenario (frame ++ and no consultation treatments).

Scenario 3: Election

“For several years now, elections in a European country have been marked by a significant abstention problem, considered by many observers as a threat to democracy. Two days before the last election, for example, it was estimated that only around half of eligible voters intended to go vote. In order [**fame + +: to increase the participation rate / fame - -: to decrease the abstention rate**], the Minister of Interior has decided [**consultation = 1: after consultation with citizen representatives**], to introduce a policy of mobilization that would not limit the freedom of voters. Thus, on the day before the elections, voters received the following SMS: “Come and vote this year: we are expecting to see a much higher turnout than in previous years!” This type of reminder combined with a message expressing a positive social reinforcement or conformism has been shown to have a greater effect than a negative message. Thus, thanks to the implementation of this policy, the polling stations registered [**fame + +: 60% participation, this is 10% more than forecasts / fame - -: 40% abstention, this is 10% less than the forecasts.**]”

Depuis plusieurs années, les élections d'un pays d'Europe sont marquées par un important problème d'abstention, considéré par beaucoup d'observateurs comme une menace pour la démocratie. Deux jours avant les dernières élections, il a été par exemple estimé qu'environ la moitié des électeurs se déplacera pour aller voter.

Afin d'augmenter le taux de participation, le Ministre de l'Intérieur a décidé de mettre en place une politique de mobilisation qui ne nuirait pas à la liberté des électeurs.

Ainsi, la veille des élections, les électeurs ont reçu le SMS suivant :

"Venez tous voter cette année : nous allons observer un taux de participation nettement supérieur aux années précédentes !".

En effet, il a été prouvé qu'un rappel associé à un renforcement ou à un conformisme positif à l'aide d'un message avait un plus grand effet qu'un message négatif.

Ainsi, grâce à la mise en place de cette politique, les bureaux de vote ont pu enregistrer 60% de participation, soit 10% en plus que les prévisions.

Figure A3: Screenshot of the election scenario (frame ++ and no consultation treatments).

Scenario 4: Wasted time at work

“In a consulting company where employees spend their days at a desktop computer, it was noticed that most of the employees spent time on social media or on websites that were not related to their work, thus diminishing the company’s productivity. In order to encourage its employees to dedicate **[fame + +: more time in the day to work-related activities / fame - -: less time in the day to non-work-related activities]** the management has decided **[consultation = 1: on the advice of the works council]** to introduce a policy of accountability which would not affect the freedom of its employees. Thus, a non-work time tracker was installed on their computer screens. This time tracker appears at the bottom right of the taskbar when they are viewing pages that an algorithm deems unrelated to their work. The tracker is programmed to keep browsing information encrypted and anonymous, so that management has no way of accessing it. This was implemented because it has been proven that this type of intervention allows employees to be more aware of their behaviour and adapt it in the desired direction. Thus, thanks to the implementation of this policy, work time dedicated **[fame + +: to work-related activities increased by 30 minutes per day / fame - -: to non-work-related activities decreased by 30 minutes per day.]”**

Dans une entreprise de conseil où les salariés passent leurs journées sur un ordinateur de bureau. Il a été remarqué que la plupart des employés passaient du temps sur les réseaux sociaux ou sur des sites internet qui n’étaient pas en lien avec leur travail, diminuant ainsi les résultats de l’entreprise.

Afin d’inciter les salariés à dédier plus de temps dans leur journée aux activités qui concernent leur travail, la direction a décidé de mettre en place une politique de responsabilisation qui ne nuirait pas à la liberté de ses employés.

Ainsi, un compteur d’heures non-travaillées a été installé sur les ordinateurs. Ce compteur s’affiche en bas à droite de la barre des tâches, uniquement lorsque vous visualisez des pages qu’un algorithme juge ne pas être en lien avec votre travail. Le compteur est programmé de façon à ce que vos informations de navigation soient cryptées et restent anonymes, de manière à ce que votre direction n’ait aucun moyen d’y avoir accès.

En effet, il a été prouvé que ce type d’intervention permettait aux salariés de mieux avoir conscience et d’adapter leur comportement dans le sens désiré.

Ainsi, grâce à la mise en place de cette politique, le temps moyen qu’un employé dédie aux tâches en lien avec son travail a augmenté de 30 minutes par jour.

Figure A4: Screenshot of the company scenario (frame ++ and no consultation treatments).

Appendix 2: Acceptability Scale

Merci d'évaluer la politique mise en place.

	Pas du tout d'accord	Plutôt en désaccord	Ni d'accord, ni en désaccord	Plutôt d'accord	Tout à fait d'accord
Soutenez-vous cette politique ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Vous opposez-vous à cette politique ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trouvez-vous cette politique éthique ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trouvez-vous cette politique manipulatrice ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trouvez-vous cette politique immorale ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trouvez-vous cette politique coercitive ?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure A5: Screenshot of the acceptability scale (in French)