

Allies do not back down
A cross-national survey experiment on audience costs
among the European public¹

Pierangelo Isernia, Sergio Martini, Francesco Olmastroni²

Department of Social, Political and Cognitive Sciences

University of Siena

Abstract

The theory of audience costs argues that national publics punish their leaders when the latter threaten the use of force in a foreign policy crisis and then back down as citizens would penalize inconsistency between words and deeds. Yet, its implications have been generally tested in the case of the United States, with basically no research on either the European Union or its member states in spite of the new initiatives launched to deepen the EU's Common Security and Defence Policy and the increasing need for a coordinated defense after recent international crises (e.g., Afghanistan, Ukraine). We contribute by reporting on a set of randomized online survey experiments with representative samples in France, Germany, Italy and the United Kingdom. In particular, we analyze how the European publics react when exposed to a possible foreign policy crisis and what effects inconsistent governments' behaviors have on public approval also depending on the stance taken by other European allies. This article presents the first large-scale cross-national experiment testing the audience cost theory in European parliamentary democracies, with important implications for the study of European citizens' preferences on foreign military crises and the enactment of a common defense strategy.

Keywords: Audience costs; Common Security and Defense Policy; Military intervention; Costs of war; Survey experiments; Europe.

¹ The project is funded by the Volkswagen Foundation, under the "European Common Defence and Shared Security in an Age of Brexit and Trump" project (Grant Letter Ref. 94760, see: <http://www.security.eu/>).

² Corresponding author: olmastroni3@unisi.it

Introduction

Domestic political conditions shape foreign policy decisions in democratic systems. According to the audience cost theory, leaders incur in reputation costs if they escalate a foreign military crisis through the threat of the use of force and then back down (Fearon 1994, 1997; Schultz 2001; Slantchev 2006; Smith 1998). The possibility of suffering electoral losses and drops in approval ratings would prevent democratic leaders from issuing empty threats, making their international commitments more credible than those of their autocratic counterparts (Gelpi and Griesdorf 2001; Clare 2007; *contra* Weeks 2008). The theory has been developed considering almost always the case of the United States (US), with basically no research on the European Union (EU) and its member states (e.g., Thomson 2016). This is largely justified, since the US is a world leader and a security provider while the EU still lacks common infrastructures for military intervention and defense, with each member state deciding to engage or not in foreign military crises as the result of international agreements and its participation in supranational institutions (NATO, UN).

In recent years, however, new initiatives have been launched to deepen the EU's Common Security and Defence Policy (CSDP), including the release of a Global Strategy on Foreign and Security Policy in 2016, the awakening of the Permanent Structured Cooperation (Juncker 2017) and the adoption of the Strategic Compass for Security and Defence in March 2022. This flurry of activity occurred against the backdrop of difficult challenges facing the continent. New threats to European security, from the resurgence of Russia to the dangers of international terrorism, and recent international crises (e.g., Afghanistan, Ukraine) urge for a coordinated approach to defense policy and feed the expectation of a common

EU response among the national publics. Interestingly, this expectation is accompanied by stable public support for CSDP throughout Europe as well as the preference for a more independent defense cooperation (Mader et al. 2021).

Can leaders and governments generate audience costs in parliamentary democracies where policy responsibility is more diffused as in the case of Europe? Do European citizens punish the behavior of leaders and governments as a result of their inconsistencies in a foreign policy intervention? And, finally, do the European publics react differently depending on how the other allies in the continent behave in a military crisis? We try to address these questions and study audience costs in Europe, reporting on a set of pre-registered randomized online survey experiments conducted on representative samples in France, Germany, Italy and the United Kingdom (UK) (Isernia et al. 2021). In the experiments, respondents were asked about their evaluation of their own government's performance during a hypothetical foreign military crisis in which two conditions varied in a fully factorial design: 1) the *type of commitment* made by the government, and 2) the *behavior of other European allies*.

This way the article contributes to the literature on audience costs in international crises as well as to the study of foreign and defense policy and the role of public opinion in Europe. First, it provides comparative findings for the assumptions advanced by the audience cost theory showing that the costs generated by inconsistent leaders' behaviors extend to European parliamentary democracies. However, how the public reacts seems to depend on the stances taken by other European allies, though inconsistency costs and the effect of alliances vary according to the national context. This has more general implications for expanding our understanding of the preferences of European citizens on foreign military crises and the enactment of a common defense strategy as the European publics may consider the role of European alliances and the need for coordination when intervening in an international dispute.

Theoretical framework

Audience costs in a military intervention

According to pioneering work on international bargaining (Schelling 1960), one way for leaders and governments to signal their intention in foreign policy crises and change the opponent's behavior is the generation of audience costs. This may be pursued by threatening the use of force since citizens tend to punish inconsistency between words and deeds in politics (Fearon 1994). Thus, the costs might take the form, for instance, of rising disapproval or electoral punishment of incumbents, and they are usually motivated by concerns over the credibility and reputation of the nation, that is to say, whether other countries trust public commitments made by the country's leadership during the early and ongoing stages of a foreign policy crisis. Leaders' inconsistent behaviors are, in fact, generally considered as a sign of lack of competence with a consequent loss of status compared to other national powers (Guisinger and Smith 2002; Schultz 2001). The audience cost would thus be the increase in disapproval that occurs whenever a leader makes commitments and does not follow through (Tomz 2007).

In spite of increasing attention toward the audience cost theory, quantitative tests of its main implications via observational data still rely on mixed results. Some studies have highlighted that democratic states generate higher audience costs vis-à-vis autocratic states (Gelpi and Griesdorf 2001; Schultz 2001). Other research has suggested that the capacity of democracies to generate audience costs may hinge on the public's access to mass media communication, the format of electoral institutions structuring competition among parties and the relationship between government and opposition (Potter and Baum 2013; Weeks 2008).

Observational studies have also raised important criticism (Gartzke and Lupu 2012). One problem is that international outcomes are the results of strategic

choices of individual actors. It is therefore likely to observe cases in which presidents back down from previous threats only when the costs of inconsistency are expected to be very low (Schultz 2001). This implies that any inference may be influenced by selection processes. Following up on this, a second problem is that large-N studies based on observational data include instances that are not actually effective threats (Downes and Sechser 2012). Last, audience costs experienced by democratic leaders are not necessarily comparable with those of their autocratic counterparts as their magnitude may depend on other non-observable characteristics related to the type of regime, leading therefore to an endogeneity bias in obtained results.

Experiments, on their part, may be better suited to overcome these drawbacks of observational research. Indeed, experimental studies have become increasingly influential in the field, as they profit from a strict control of the research environment and from the randomization of treatments. In this respect, experimental research on audience costs has found that citizens tend to punish inconsistent behaviors by their leaders and that the size of such costs may be influenced by specific features of the international environment, such as the type of regime toward which the threat is oriented or the level of escalation (Tomz 2007). Other studies have, instead, looked at how US domestic politics, in the form of the justification the president provides for backing down, the president's partisanship and whether parties in Congress support his decision (Levendusky and Horowitz 2011; Trager and Vavreck 2011; Evers et al. 2019), reduces or amplifies public punishment. Moreover, a few instances have shown that audience costs not only arise when leaders fail to implement threats, but also when they fail to honor promises to stay out of a conflict (Levi et al. 2015). Last, it has been emphasized that individual predispositions toward the use of military force and the distinction between 'hawks' and 'doves' moderate citizens' approval of different policy decisions during a foreign policy crisis (Kertzer and Brutger 2016).

Overall, with the exception of a few studies that have run similar experiments in the UK (Davies and Johns 2013; Thomson et al. 2016), no available research provides a large-scale cross-national experiment to test the audience cost theory in European parliamentary democracies.

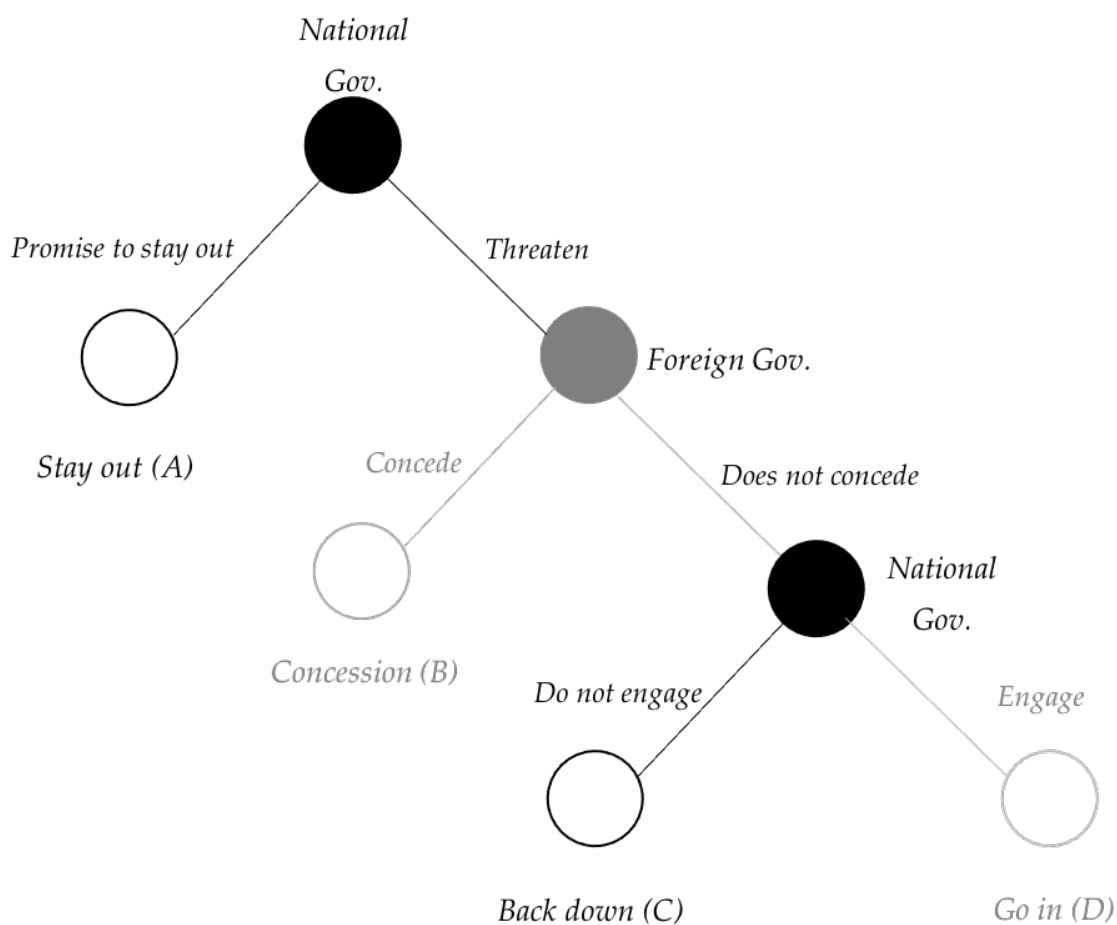
Audience costs in Western European democracies

Audience costs theory evaluates predictions based on standard crisis bargaining games with two players, a sender and a target (Fearon 1994). For the sake of simplicity, we could employ a national European government as the former and any foreign challenger government as the latter. In this framework, the model assumes the two countries fighting over a certain good, with some costs related to any specific decision and course of action. If the two countries start the conflict, the one that wins the war obtains the expected good. It is, thus, possible to expect people to have different preferences depending on these costs.

As illustrated in Figure 1, this scenario can be represented as a series of sequential moves and decision nodes involving the two players. The game starts with the sender deciding whether to threaten the target in response to an undesirable action (i.e., any act that might have endangered interstate security, such as the military invasion of a third country). If the sender “promises to stay out”, then the game ends with the sender being out of the conflict, that is condition A. On the contrary, if it decides to “threaten” the target state, then, the target might concede leading to condition B or continue with its aggressive operations. If the target state does not concede, next, the sender might follow through on its threat and engage in a war against its opponent, namely outcome D, or back down without engaging in the conflict, leading to scenario C. The literature considers any existing difference in terms of public approval between the ‘stay out’ condition A and the ‘back down’ condition C as the audience cost for not being consistent

with the initial commitment. In fact, according to the theory, citizens are expected to prefer a situation in which a government refrains from entering a conflict without making any threat than a situation in which the executive exposes the country to a loss in reputation, issuing 'empty threats' (Fearon 1994; see also Schultz 2001; Smith 1998; Tomz 2007).

Figure 1: Crisis bargaining model



Our study adds to the literature exploring some aspects so far neglected by the research on audience costs. First, it is important to apply it to the European context as the mechanisms underlying audience costs generated by leaders' failure to comply with prior commitments might be different in semi-presidential and

parliamentary democracies as compared to Presidential systems. To this purpose, we test the audience cost theory in four Western European democracies, namely France, Germany, Italy and the UK. While it should be acknowledged that other political figures and officials are involved in the foreign policy decision-making process and that a multiplicity of domestic actors enter this process with their preferences, governments, including both premiers and key cabinet ministers (i.e., defense and foreign affairs), are generally considered the traditional architects of a country's foreign and defense policy. They participate in person in the framing of foreign policy and security issues and, as a result, they are deemed ultimately responsible for foreign policy decisions by their own publics (Olmas-troni 2014). Moreover, a favorable public rating of the government's performance is likely to be precondition for electoral success of the ruling parties. Hence, it becomes important to test whether executives are likely to experience audience costs in parliamentary democracies as a result of their false commitments in foreign policy. Given these premises and in line with what previous literature on the US case suggests, we advance the following hypothesis:

H1: We expect disapproval of a government's conduct to be higher when it makes a threat and then backs down of the conflict compared to when the government decides to stay out without issuing any threat. So, $dis(C) > dis(A)$.

Second, when evaluating the relevance and impact of audience costs it is important to examine the role of other allies (Gaubatz 1996; Smith 1996). Military alliances, namely an agreement established with other nations with respect to coordinated intervention in case of military confrontation, might affect the costs of inconsistency between words and deeds, muting or amplifying audience costs depending on whether the behavior of the national government matches that of its allies.

This might be due to several mechanisms (Smith 1996; Tomz and Weeks 2021). In the case of empty threats, that is, when the national government initially threatens the use of force and then back down, not intervening in a military conflict, the cost generated might be magnified when the allies decide to participate in the operation, so that the national government not only diverts from its declared stance but also de-aligns from its partners. This instead might be reduced when the allies decide to not participate, the same way as the national government. Alliances, in fact, build on the expectation of successful cooperation, so breaking previous agreements affects reputation for military reliability as well as for non-military reliability like, for instance, trade or immigration, all relevant issues in the European context (Crescenzi et al. 2012; Downs and Jones 2002). Thus, we expect:

H2: Audience costs to be lower (higher) if the behavior of the national government is in line with (contrary to) that of other European allies. So, $dis(C_{Allies\ in}) - dis(A_{Allies\ in}) > dis(C_{Allies\ out}) - dis(A_{Allies\ out})$.

Data and Experimental Design

We evaluate our arguments using a set of survey experiments fielded online in Italy in October - November 2021 and in France, Germany and the UK in December 2022. Having settled the study and collected data before the outbreak of the Russian-Ukrainian conflict and some months after the withdrawal of the US troops from Afghanistan allowed us to elicit citizens' reactions in a period in which the European public debate was less polarized and the topic of foreign policy and the use of military force was less salient than at the time of writing.

Our experimental design builds on the general approach developed in Tomz's seminal contribution (Tomz 2007) and constituting a benchmark for many other

studies on audience costs (e.g., Levendusky and Horowitz 2011; Trager and Vavreck 2011; Levy et al 2015; Kertzer and Brutger 2016; Evers et al. 2019). In order to make our results comparable with those of previous research on the same topic, we employed a similar setup, avoiding any language that might have been open to interpretation, and thus bias our results either favoring an effect of inconsistent behavior on governmental approval or inducing a moderating role for alliances. As discussed below, however, we introduce some relevant adjustments in order to evaluate our arguments. The manipulations we present and analyze below represent only a fraction of the total number of treatments assigned in the study as we focus on those useful to test our hypotheses (Isernia et al. 2021).

Respondents began the survey reading a short introduction about a situation the specific European country (being this either France, Germany, Italy or the UK) “might face in the future” and the fact that “different leaders might handle the same situation in different ways”. Then, respondents were informed that they were going to be asked to evaluate one approach that their country’s government might take and ask whether they approved or disapproved it. Subjects were randomly assigned to one of four different experimental groups resulting from the combination of two treatments. Table 1 summarizes all the experimental conditions and lists the number of respondents assigned to each group (see the Appendix for the full experimental protocol).

To check for the robustness of the random assignment, we performed some balance tests (Hansen and Bowers 2008) by multinomial regression, conditioning the assignment to each experimental group on a set of predictors ranging from socio-demographic characteristics (i.e., gender, age, education) to ideology. Results confirm that the random procedure was correct in the four countries, with no variable being statistically different across the treatment groups. Pseudo-R squared terms are always very small and the likelihood ratio chi-squared tests are

not significant (see Figure A1-A4 in the Appendix). Hence, any difference between conditions should be attributed to treatment manipulation only and not to other confounding factors.

Table 1: Randomly assigned conditions in the experiments

France	Commitment	European allies	N
A ₁	Promise to stay out	Decided not to participate	181
A ₂	Promise to stay out	Decided to participate	170
C ₁	Threat use of force	Decided not to participate	150
C ₂	Threat use of force	Decided to participate	163
Germany	Commitment	European allies	N
A ₁	Promise to stay out	Decided not to participate	158
A ₂	Promise to stay out	Decided to participate	188
C ₁	Threat use of force	Decided not to participate	174
C ₂	Threat use of force	Decided to participate	186
Italy	Commitment	European allies	N
A ₁	Promise to stay out	Decided not to participate	240
A ₂	Promise to stay out	Decided to participate	252
C ₁	Threat use of force	Decided not to participate	261
C ₂	Threat use of force	Decided to participate	255
United Kingdom	Commitment	European allies	N
A ₁	Promise to stay out	Decided not to participate	182
A ₂	Promise to stay out	Decided to participate	179
C ₁	Threat use of force	Decided not to participate	171
C ₂	Threat use of force	Decided to participate	174

Each hypothetical scenario presented the respondents with a crisis in which an unspecified foreign country used military power against another unnamed neighboring country. Unlike other research (Thomson 2016), we decided to not specify the attacking and the attacked country's names to avoid any contamination of answers with respondents' prior opinions about real entities. Our design differs from previous research since we try to assess audience costs and the moderating effects of the stance taken by allies in the European context. The scenario involves a military conflict in which "a country sent its military to take over a neighboring country", a situation that well resembles the recent Russian-Ukrainian conflict despite the experiments being fielded before the outbreak of this crisis. Here, the option for the national government is whether or not to use military force against the attacking country. Overall, we manipulated two conditions giving us a 2x2 fully factorial design.

The first treatment considered whether or not the national government reacted to an aggressive behavior against a third country (i.e., a military invasion) by admonishing the attacking state that the country's military "would push out the invaders". The alternative was that the government promised to "stay out of the conflict". Differently from Tomz (2007), we neither varied the level of escalation in the threat on the part of the national government, nor manipulated other aspects, such as the type of regime of the attacking country, the motives that drove the attack, the military capability of the invader, or the level of threat posed by the invasion to the European country's national security.

The second treatment manipulated the behavior of other European countries notifying whether these had aligned or not with the national government, so that the respondent was informed that "other European allies have decided to participate" or "to not participate" in the military intervention. This allowed us to explore whether the cost of being inconsistent with respect to a previous threat to

the use of force to solve an international dispute is moderated by what other European alliances do, and whether the public in each country values multilateral intervention and coordination. In this respect, our design differs from other studies in which the role of alliances is explored when the ally is the victim of aggression (Davies and Johns 2013; Tomz and Weeks 2021).

After being told that “the attacking country continued to invade”, the respondent was informed that, as a response, the national government decided to “not engage the military”. This allows us to assess the impact of following through or backing down with respect to a bargaining action. Finally, at the end of the scenario the respondents had to express their approval of the national government’s conduct by answering a question whose format largely resembles those already employed in previous studies (e.g., Tomz 2007; Levendusky and Horowitz 2012; Levy et al. 2015). Answers were collected using a 7-point Likert scale ranging from ‘strongly approve’ to ‘strongly disapprove’.

The French, German, and British surveys were fielded by YouGov using nation-specific online open-access panels. The respondents for the surveys were recruited based on quotas on gender, age-group and educational attainment. In Italy, the survey was fielded by GfK. In this case, the sample was selected within a probability panel held and managed by the same company. GfK recruits subjects regardless of their access to the Internet. If they accept to be part of the panel, they are, then, provided with a tablet to participate in incentivized online surveys. In the first stage, households are selected via a stratified random sampling using region, urban area and number of household members as main strata. In the second stage, individuals are then randomly selected according to gender, age, geographical area and the size of urban area quotas. In the research we surveyed 6,300 respondents in total (FRA: 1,409; GER: 1,458; ITA: 2,049; UK: 1,384), though we focus on a random subsample of 3,084 to whom the experimental

treatments under examination were administered (FRA: 664; GER: 706; ITA: 1,008; UK: 706).

Empirical analysis

Since our study aims to evaluate audience costs suffered due to inconsistent decision-making, our dependent variable measures the respondent's level of disapproval of the conduct maintained by her/his government through an ordinal 7-point scale. To identify the effects of our treatments, we estimate a series of ordered logistic regression models (Long 1997). In an ordinal logit, the observed dependent variable y_i is assumed to be associated with an underlying continuous, latent variable y_i^* , according to the following measurement equation:

$$y_i = c \quad \text{if} \quad \alpha_{c-1} < y_i^* < \alpha_c \quad \text{for} \quad c = 1, \dots, C$$

The latent trait y_i^* underlying the level of disapproval for the i -th individual goes from $-\infty$ to $+\infty$ and it is measured using the observed variable y_i , which is our indicator of disapproval of the way the government handled the situation, with α representing the cut points or thresholds. Our dependent variable has seven categories C , going from 1, 'strongly approve', to 7, 'strongly disapprove', so that the relationship between latent and observed variable is determined by:

$$y_i \begin{cases} 1 \Rightarrow \text{Strongly approve} & \text{if } -\infty \leq y_i^* < \alpha_1 \\ 2 \Rightarrow \text{Approve} & \text{if } \alpha_1 \leq y_i^* < \alpha_2 \\ 3 \Rightarrow \text{Tend to approve} & \text{if } \alpha_2 \leq y_i^* < \alpha_3 \\ 4 \Rightarrow \text{Neither / Nor} & \text{if } \alpha_3 \leq y_i^* < \alpha_4 \\ 5 \Rightarrow \text{Tend disapprove} & \text{if } \alpha_4 \leq y_i^* < \alpha_5 \\ 6 \Rightarrow \text{Disapprove} & \text{if } \alpha_5 \leq y_i^* < \alpha_6 \\ 7 \Rightarrow \text{Strongly disapprove} & \text{if } \alpha_6 \leq y_i^* < \infty \end{cases}$$

When falling into the category, the equation is:

$$P(y_i = c|x) = \text{Logit}^{-1}(\alpha_c - X_i\beta + \varepsilon_i)$$

where we model the probability of expressing different degrees of disapproval in each observation i . This is linked via the inverse of the *logit* to α_c , that is, the cut point for category c , X_i , the row vector of data for the i -th observation, β , the column vector of coefficients, and ε_i , the error term which has a logistic distribution. In our case, β is the combination of the different treatments in the form of dummy variables capturing the effects of the type of commitment, eliciting the consistency of the decision and the position held by the European allies as well as their interactions.³

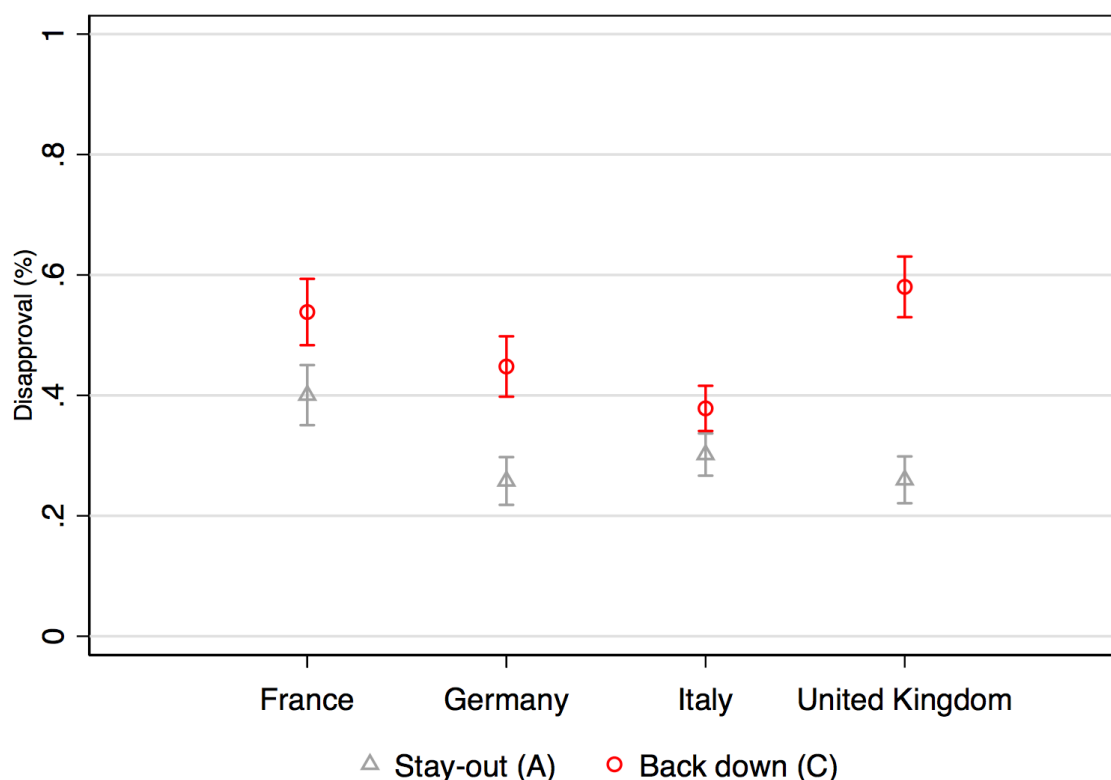
For the sake of simplicity, we mainly display graphical results showing predicted probabilities of negative outcomes, that is, the extent to which respondents tended to disapprove, disapproved or strongly disapproved as a function of our covariate. To test differences in the effect of treatments, we computed average marginal effects and performed Wald tests (full results are shown in Table A1 in the Appendix).⁴

Figure 2 presents all the relevant quantities useful to compute the audience costs of issuing empty threats across the four examined countries, specifically: the percentage of each sample that expressed disapproval in the “Stay out” condition and the percentage that expressed disapproval in the “Back down” condition C.

³ Experiments rely on the manipulation of the variables of interest and their random assignment, which ensure overcoming the problem of endogeneity and confounding variables by a strict control of the design. This allows avoiding the inclusion of control variables (e.g., socio-demographic factors).

⁴ In the Appendix we also include more easy-to-interpret cross-tabulations for the main outcomes of interests, whose results do not diverge from those presented in the manuscript (see Tables A2, A3A and A3B).

Figure 2: Audience costs of empty threats

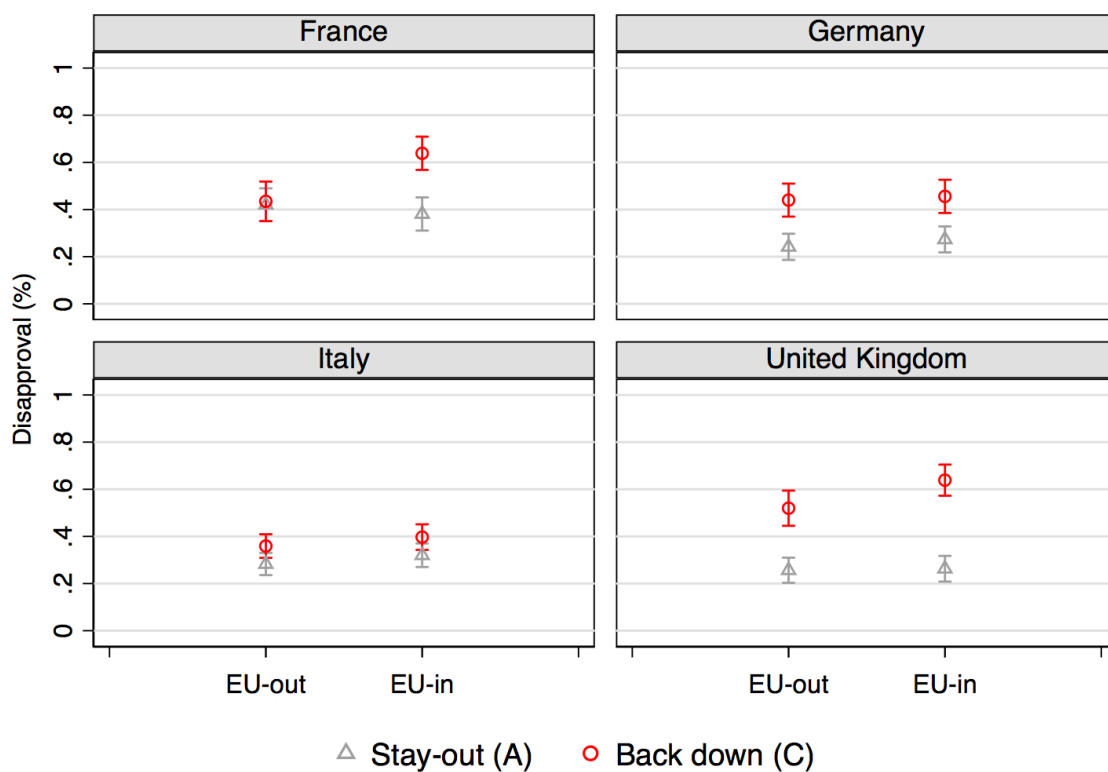


Note: lines around shapes represent 95% confidence intervals.

As it can be clearly seen, citizens tend to punish inconsistent governments. The level of disapproval is, in fact, fairly higher when the government backs down after an initial threat than when it maintains an initial promise not to intervene. This holds true in all the considered national contexts, such that the differences in proportions are statistically significant ($p < 0.01$ in all the cases). However, we find important country variations. We find the largest effect in the UK with an audience cost for empty threats equal to a 32-percentage point increase in disapproval. This is almost two times the audience cost we find in Germany (17%), almost three times that of France (12%), and more than four times the value obtained in Italy (8%). In sum, these results provide support for our first hypothesis

(H1) and confirm the assumptions of the audience cost theory also in the European context. Indeed, the rise of disapproval is noticeable and, in some circumstances, larger than what it is reported in other studies of the US context (e.g., Tomz 2007; Trager and Vavreck 2011).

Figure 3: Audience costs of empty threats depending on European allies' position



Note: lines around shapes represent 95% confidence intervals.

Moving to our second expectation, in which we aim to test the idea that audience costs hinge on contextual factors such as the behavior of other allies, we can assess it by considering the interaction between the commitment made by the national government and the stance taken on the conflict by the European allies. In this respect, Figure 3 partitions audience costs according to whether other

European countries decided to participate or not in the conflict. As it can be noticed, the effect especially holds true in France where the interaction is statistically significant ($p < 0.01$). Indeed, when European allies take part in the conflict the costs of backing down from an initial threat further increase by about 11 percentage points (23%, $p < 0.01$) as compared to the previous findings for this country (12%, $p < 0.01$). By contrast, when European allies do not take part in the conflict, the cost of backing down is basically reduced to zero in France.

Similarly, in the UK we see an increase in audience costs when the government de-aligns from the other European allies; however, this effect is lower in size (38% $p < 0.01$; +6%). Conversely, when there is an alignment between the national government and the European allies, audience costs decrease to the same extent (26% $p < 0.01$; -6%). All in all, despite a change in audience costs depending on the alliances is observable in the UK, this is not strong enough to get a statistically significant interaction effect (similarly the delta between the audience cost in the "EU-out" condition and in the "EU-in" condition is barely significant: +10%, $p < 0.08$). We conclude our examination, considering Germany and Italy. As it can be seen, in both countries there is no interaction between audience costs and the posture of the European allies, as the effect of alliances is negligible.

Conclusion

In recent years, the audience cost theory has considerably attracted the attention of international relations scholars as it integrates the constructivist approach to foreign policy decision-making with rational choice models of crisis bargaining. The conventional logic of this theory is that democratic governments would incur in domestic audience costs when they publicly expose themselves to cheap talks and false commitments. In spite of the increasing attention, there is much debate on whether audience costs actually exist and under what circumstances we can

observe them. Experimental research has proved to be better equipped to avoid some basic identification problems that are instead common with observational data, pointing at the importance of factors related to the international environment as well as to the domestic political arena.

This article has proposed a new analytical framework to account for audience costs in different democratic contexts, with multiple decision-making patterns and a new condition so far neglected but of great importance in the European context and in view of a further integration in defense and foreign affairs, that is, the contribution offered by allied countries during a military confrontation with a third country. In this respect, following seminal research, we have examined audience costs in Europe, testing whether governments suffer disapproval from their publics when they issue empty threats about the use of force. Moreover, we assessed the difference between reneging on a public threat depending on whether or not other European countries took part in the conflict. Here, the rationale is that citizens would punish their governments less if the latter align their behavior to that of their international partners.

By means of new and original randomized online survey experiments conducted in France, Germany, Italy and the UK, we have explored all these aspects showing that audience costs related to empty threats are not limited to the US context, but they also apply to European democracies. Thus, mechanisms of executives' punishment are common to different institutional contexts.

Our analysis has also shown that the position taken by other European allies may moderate the cost generated by governments' inconsistent behaviors, suggesting that backing down from an initial threat may result in further disapproval of the national government if this implies a dealignment from the international partners. Interestingly, this result is particularly relevant for those European countries that are traditionally considered as more interventionist (i.e., France and, in the same direction but with a lesser precise effect and estimate, in the UK) while the

effect is negligible in the case of the other two middle powers (Santoro 1991) under examination (Germany and Italy). On the contrary, similar country patterns may be observed when the European allies decide to not take part in the conflict. In this case, the costs of inconsistency are reduced, suggesting the importance of a multilateral approach to crisis management and dispute resolution in the European context.

All in all, this article has contributed to the audience cost literature in two different ways. First, it has accounted for its generalizability, testing whether the theory holds varying the institutional context under consideration. By providing evidence that audience costs also arise in European democracies, the article improves our understanding of the domestic prize of inconsistent behaviors in crisis bargaining. Second, our research has demonstrated that in the European context the magnitude of audience costs may also be dependent on the position held by the allies. These findings confirm that citizens care about both the process of foreign decision-making, although with an important division among more and less traditionally interventionist countries. Overall, it has to be noticed that we find important differences across countries both with respect to the observed audience costs as well as to the moderating role of alliances.

This has both substantive implications for theories of signaling in international affairs and policy implications for the development of a CSDP. Since citizens tend to punish ineffective governments who break prior commitments to become involved in conflicts, especially if this occurs without the participation of a large European coalition, executives wishing to preserve their reputation should avoid issuing an empty threat if other European partners are not likely to engage. These results also suggest that a multilateral approach to foreign and defense policy is not only supported by majorities of Europeans (Mader et al. 2021), but it is also a precondition to avoid public disapproval or electoral punishment at the occurrence of a military crisis.

References

- Clare, Joe. 2007. "Domestic Audiences and Strategic Interests." *Journal of Politics*, 69 (3): 732–45.
- Crescenzi, Mark, Jacob Kathman, Katja Kleinberg, and Reed Wood. 2012. "Reliability, Reputation, and Alliance Formation." *International Studies Quarterly* 56: 259–274.
- Davies, Graeme A. M. and Robert Johns. 2013. Audience Costs among the British Public: The Impact of Escalation, Crisis Type, and Prime Ministerial Rhetoric, *International Studies Quarterly* 57 (4): 725–737.
- Downs, George w., and Jones, Michael A. 2002. "Reputation, Compliance, and International Law." *Journal of Legal Studies* 31: S95–114.
- Downes, Alexander B., and Todd S. Sechser. 2012. "The Illusion of Democratic Credibility". *International Organization*, 66: 457–89.
- Eyerman, Joe, and Robert A. Hart, Jr. 1996. "An Empirical Test of the Audience Cost Proposition." *Journal of Conflict Resolution* 40 (4): 597–616.
- Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88 (3): 577–92.
- Fearon, James D. 1997. "Signaling foreign policy interests: Tying hands versus sinking costs." *Journal of Conflict Resolution* 41 (1): 68–90.
- Gaubatz, Kurt Taylor. 1996. "Democratic States and Commitment in International Relations." *International Organization* 50(1): 109–39. doi: 10.1017/S0020818300001685.
- Gartzke, Erik, and Yonatan Lupu. 2012. "Still Looking for Audience Costs". *Security Studies* 21: 391-97.
- Gelpi, Christopher, and Griesdorf, Michael. 2001. "Winners or Losers? Democracies in International Crisis, 1918–94". *American Political Science Review* 95: 633–48.

- Guisinger, Alexandra, and Alastair Smith. 2002. "Honest Threats: The Interaction of Reputation and Political Institutions in International Crises." *Journal of Conflict Resolution* 46 (2): 175–200.
- Hansen, Ben B, and Jake Bowers. 2008. "Covariate Balance in Simple, Stratified and Clustered Comparative Studies". *Statistical Science* 23 (2): 219–36.
- Isernia, P., Martini, S., & Olmastroni, F. (2021, October 17). SecEUrity Project: An Audience Cost Experiment in Europe. <https://doi.org/10.17605/OSF.IO/4Z3UB>
- Kertzer, Joshua D., and Ryan Britger. 2016. "Decomposing audience costs: Bringing the audience back into audience cost theory". *American Journal of Political Science* 60 (1): 234–49.
- Juncker Jean-Claude. 2017. "Speech at the Defence and Security Conference Prague: In Defence of Europe." European Commission, June 9, available at http://europa.eu/rapid/press-release_SPEECH-17-1581_en.htm.
- Levendusky, Matthew S., and Michael C. Horowitz. 2012. "When Backing Down is the Right Decision: Partisanship, New Information, and Audience Costs." *Journal of Politics* 74 (2): 323–38.
- Levy, Jack S, McKoy Michael K., Poast Paul, and Geoffrey P. R. Wallace. 2015. "Backing out or backing in? Commitment and consistency in audience costs theory." *American Journal of Political Science* 59: 988–1001.
- Long, Scott J. (1997). *Regression Models for Categorical and Limited Dependent variables*. London: Sage.
- Mader, Matthias, Francesco Olmastroni, and Pierangelo Isernia (2021) "The Polls—Trends: Public Opinion Toward European Defense Policy and Nato: Still Wanting it Both Ways?" *Public Opinion Quarterly* 84 (2): 551–582.
- Olmastroni, Francesco. 2014. *Framing War: Public Opinion and Decision-Making in Comparative Perspective*. New York and Abingdon, Oxon: Routledge.

- Potter, Philip B. K., and Matthew A. Baum. 2013. "Looking for Audience Costs in All the Wrong Places: Electoral Institutions, Media Access, and Democratic Constraint". *The Journal of Politics* 76 (1): 167–81.
- Santoro, Carlo M. (1991) *La politica estera di una media potenza. L'Italia dall'Unità ad oggi*. Bologna: Il Mulino.
- Schelling, Thomas C. 1960. *Strategy of Conflict*. Cambridge, MA: Harvard University Press.
- Schultz, Kenneth. 2001. "Looking for Audience Costs." *Journal of Conflict Resolution* 45(Feb.): 32–60.
- Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.
- Slantchev, Branislav. 2006. "Politicians, the Media, and Domestic Audience Costs." *International Studies Quarterly* 50(2): 445–77.
- Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92 (3): 623–38.
- Thomson, Catarina P. 2016. "Public Support for Economic and Military Coercion and Audience Costs". *The British Journal of Politics and International Relations* 18 (2): 407–21.
- Tomz, Michael. 2007. "Domestic Audience Costs in International Relations: An Experimental Approach." *International Organization* 61 (4): 821–40.
- Tomz, Michael, and Jessica L. P. Weeks. 2021. "Military Alliances and Public Support for War." *International Studies Quarterly* 65 (3): 811–824.
- Trager, Robert F., and Lynn Vavreck. 2011. "The Political Costs of Crisis Bargaining: Presidential Rhetoric and the Role of Party." *American Journal of Political Science* 55 (3): 526–545.
- Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62 (1): 35–64.

Appendix

An Audience Cost Experiment in Europe: Experimental Stimuli

You will read about a situation our country might face in the future. Different leaders may handle the situation in different ways. We will describe one approach that [COUNTRY's] leaders might take, and ask whether you approve or disapprove.

A country sends its military to take over a neighboring country. The [COUNTRY] government says [VARIATION 1] [VARIATION 2]. The attacking country continues to invade. In the end, the [COUNTRY] government does not engage the [COUNTRY] military.

VARIATION 1:

- 1) [COUNTRY] will stay out of the conflict,
- 2) that if the attack continues, the [COUNTRY] military will push out the invaders,

VARIATION 2:

[IF VARIATION 1 == 1 THEN EITHER...OR...]

- 1a) like other European allies that have decided not to participate in the military intervention
- 1b) even though other European allies have decided to participate in the military intervention

Question: Do you approve, disapprove, or neither approve nor disapprove of the way [COUNTRY] government handled the situation?

- (1) Strongly approve
- (2) Approve
- (3) Tend to approve
- (4) Neither approve nor disapprove
- (5) Tend to disapprove
- (6) Disapprove
- (7) Strongly disapprove

(-98) Don't know

(-99) NA

The four different scenarios (English translation)

A country sends its military to take over a neighboring country. The [COUNTRY] government says [COUNTRY] will stay out of the conflict, like other European allies that have decided not to participate in the military intervention. The attacking country continues to invade. In the end, the [COUNTRY] government does not engage the [COUNTRY] Army.

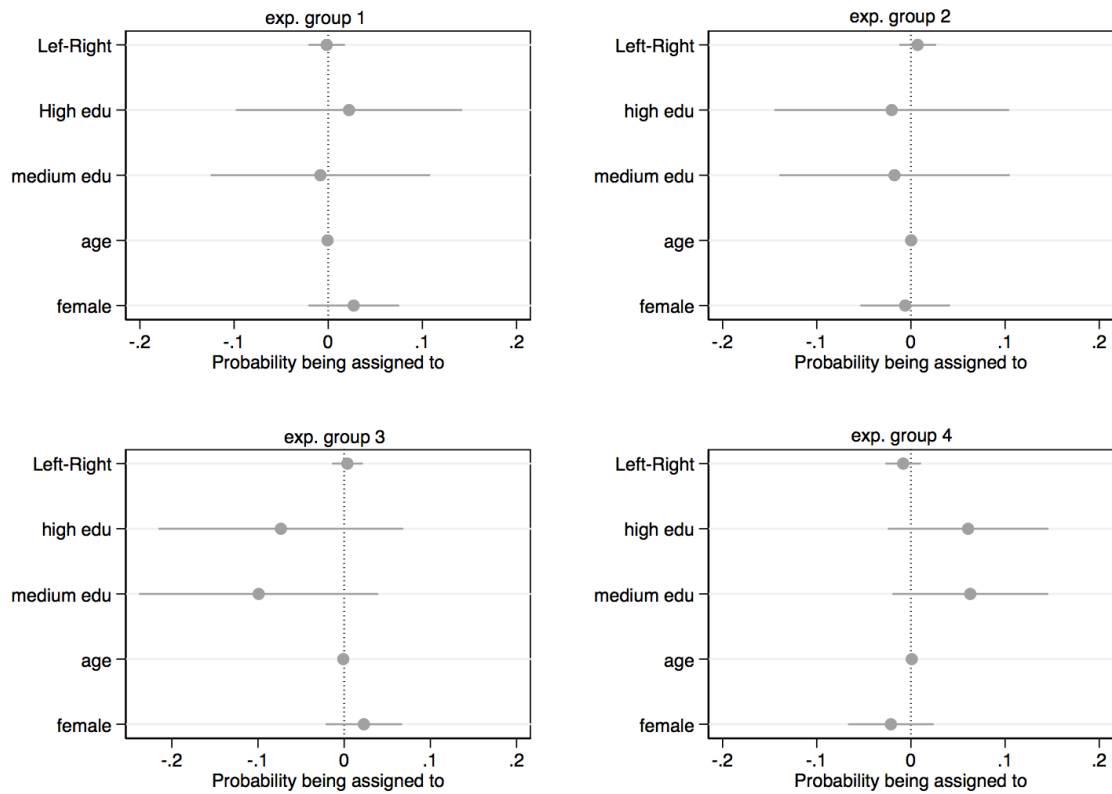
A country sends its military to take over a neighboring country. The [COUNTRY] government says [COUNTRY] will stay out of the conflict, even though other European allies have decided to participate in the military intervention. The attacking country continues to invade. In the end, the [COUNTRY] government does not engage the [COUNTRY] Army.

A country sends its military to take over a neighboring country. The [COUNTRY] government says that if the attack continues, the [COUNTRY] Army will push out the invaders, even though other European allies have decided not to participate in the military intervention. The attacking country continues to invade. In the end, the [COUNTRY] government does not engage the [COUNTRY] Army.

A country sends its military to take over a neighboring country. The [COUNTRY] government says that if the attack continues, the [COUNTRY] Army will push out the invaders, with the help of other European allies that have decided to participate in the military intervention. The attacking country continues to invade. In the end, the [COUNTRY] government does not engage the [COUNTRY] Army.

Balance tests

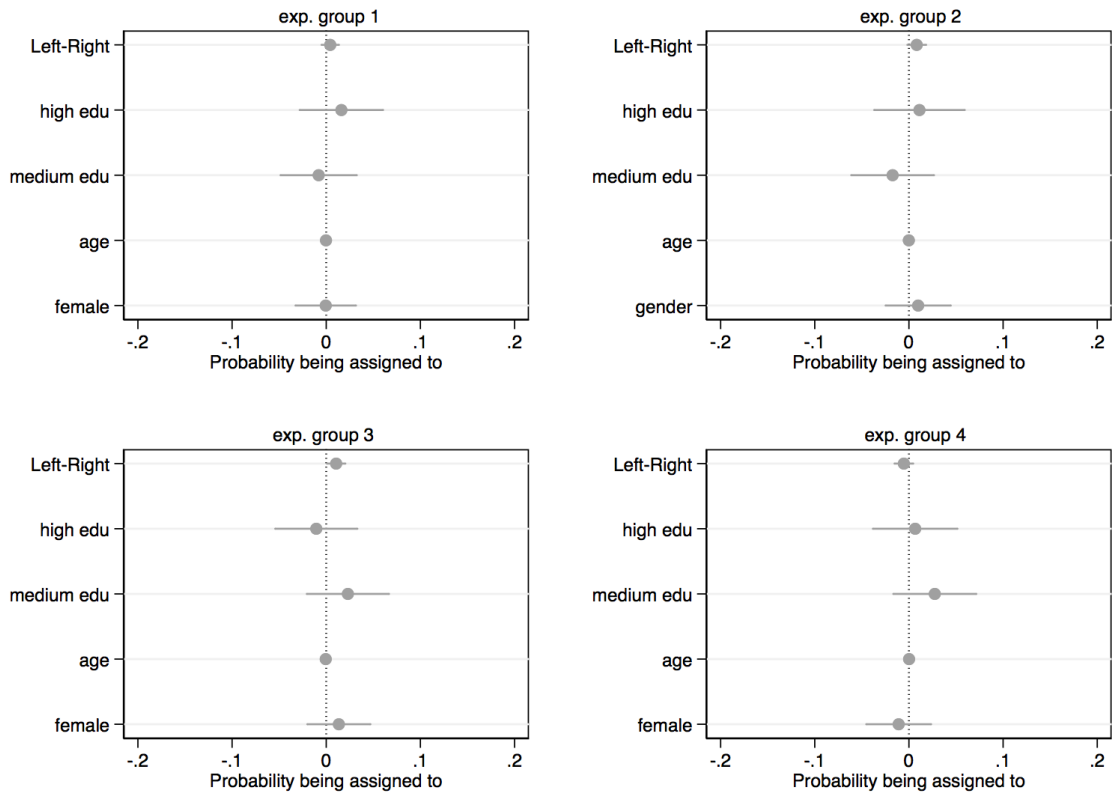
Figure A1: Balance test, multinomial logistic regression (average marginal effects) in the French sample



Note: Lines across circles represent 95% confidence intervals.

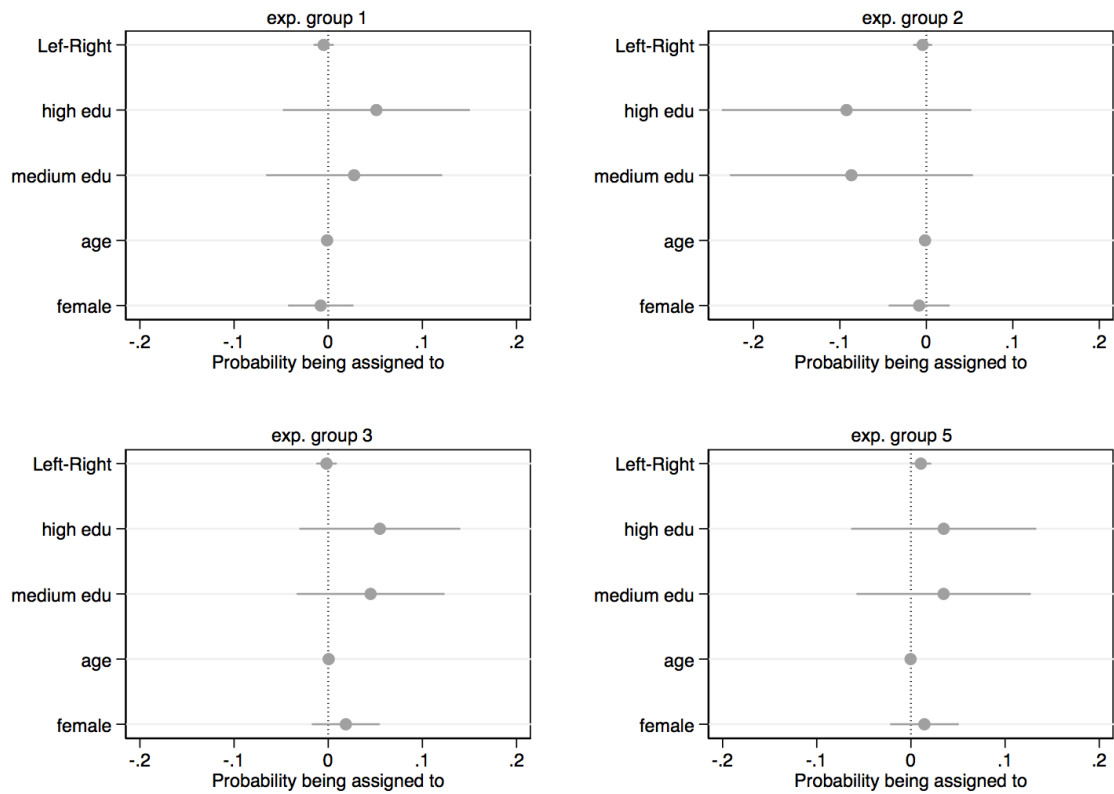
Results confirm that the random procedure was implemented safely: none of the variables increase or decrease the probability of each individual to be assigned to the experimental groups, pseudo-R squared term is very small (0.012) and the likelihood ratio chi-squared tests is not significant (0.681) (N=370).

Figure A2: Balance test, multinomial logistic regression (average marginal effects) in the German sample



Note: Lines across circles represent 95% confidence intervals. Results confirm that the random procedure was implemented safely: none of the variables increase or decrease the probability of each individual to be assigned to the experimental groups, pseudo-R squared term is very small (0.007) and the likelihood ratio chi-squared tests is not significant (0.553) (N=706).

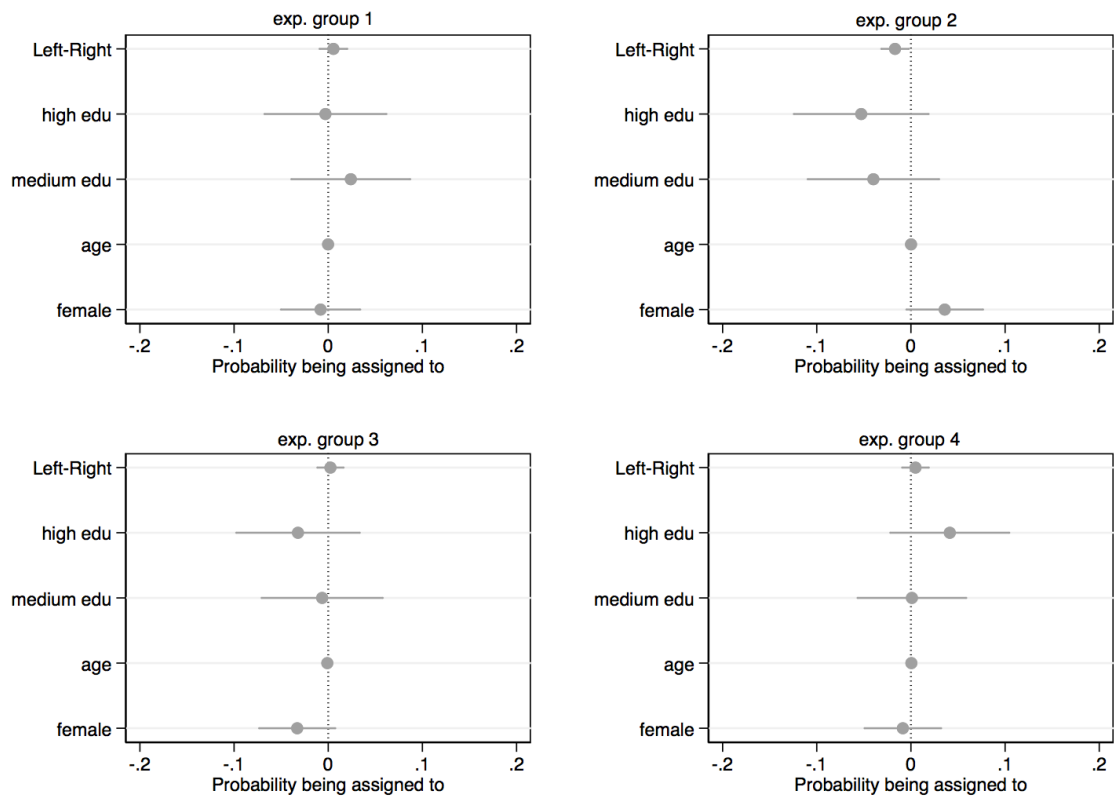
Figure A3: Balance test, multinomial logistic regression (average marginal effects) in the Italian sample



Note: Lines across circles represent 95% confidence intervals.

Results confirm that the random procedure was implemented safely: none of the variables increase or decrease the probability of each individual to be assigned to the experimental groups, pseudo-R squared term is very small (0.008) and the likelihood ratio chi-squared tests is not significant (0.483) (N=655).

Figure A4: Balance test, multinomial logistic regression (average marginal effects) in the United Kingdom sample



Note: Lines across circles represent 95% confidence intervals. Results confirm that the random procedure was implemented safely: none of the variables increase or decrease the probability of each individual to be assigned to the experimental groups, pseudo-R squared term is very small (0.013) and the likelihood ratio chi-squared tests is not significant (0.348) (N=530).

Table A1: Ordinal logistic regression (log-odds)

	France	Germany	Italy	United Kingdom
Back down	0.060 (0.217)	0.837** (0.207)	0.399* (0.164)	1.125** (0.209)
EU-in	-0.142 (0.201)	0.157 (0.201)	0.203 (0.169)	0.017 (0.199)
Back down*EU-in	0.877** (0.303)	-0.096 (0.284)	-0.012 (0.235)	0.463 (0.288)
/cut1	-2.351** (0.196)	-1.479** (0.171)	-2.515** (0.173)	-2.338** (0.210)
/cut2	-1.581** (0.165)	-0.897** (0.159)	-1.228** (0.132)	-1.008** (0.156)
/cut3	-0.627** (0.148)	-0.108 (0.153)	-0.313* (0.123)	0.039 (0.145)
/cut4	0.181 (0.145)	1.111** (0.160)	1.135** (0.129)	0.954** (0.151)
/cut5	1.238** (0.155)	2.076** (0.176)	2.144** (0.145)	2.059** (0.170)
/cut6	1.897** (0.170)	2.655** (0.192)	3.162** (0.182)	2.882** (0.192)
N	556	625	922	603
chi2	23.28	30.68	13.62	88.27

Note: Standard errors in parentheses ** p<0.01, * p<0.05.

Replication analysis: cross-tabulations

Table A2: Audience costs of empty threats

Country	% who disapproved Stay out (A)	% who disapproved Back down (C)	Costs of empty threats (C-A)
France	44 (38, 49) N=295	56 (50, 62) N=261	12** (21, 5)
Germany	27 (22, 32) N=310	43 (38, 48) N=315	16** (23, 8)
Italy	26 (22, 30) N=447	34 (30, 39) N=475	8** (13, 2)
United Kingdom	26 (22, 32) N=305	60 (55, 66) N=298	34** (40, 26)

Note: 95% confidence intervals in parentheses. * p<0.05 **p<0.01.

Table A3A: Audience costs of empty threats when European allies are in

Country	% who disapproved Stay out (A)	% who disapproved Back down (C)	Costs of empty threats (C-A)
France	41(33, 49) N=141	65 (57 72) N=141	23** (35, 12)
Germany	30 (23, 36) N=169	47 (39, 54) N=159	17** (27, 7)
Italy	28 (22, 34) N=227	38 (32, 44) N=229	10* (32, 45)
United Kingdom	30 (23, 37) N=159	67 (60, 75) N=152	38* (48, 27)

Note: 95% confidence intervals in parentheses. * p<0.05 **p<0.01.

Table A3B: Audience costs of empty threats when European allies are out.

Country	% who disapproved Stay out (A)	% who disapproved Back down (C)	Costs of empty threats (C-A)
France	46 (38, 54) N=154	47 (38, 56) N=120	1 (12, -11)
Germany	24 (17, 31) N=141	39 (31, 47) N=156	15** (25, 5)
Italy	24 (19, 30) N=220	30 (25, 36) N=246	6 (14, -2)
United Kingdom	22 (16, 44) N=146	52 (44, 60) N=146	30** (40, 19)

Note: 95% confidence intervals in parentheses. * p<0.05 **p<0.01.