UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

The Origins of Elite Persistence: Evidence from Political Purges in post-World War II France

Toke S. Aidt | Jean Lacroix | Pierre-Guillaume Meon

Abstract

This paper studies a new mechanism that allows political elites from a non-democratic regime to survive a democratic transition: connections. We document this mechanism in the transition from the Vichy regime to democracy in post-World War II France. The parliamentarians who had supported the Vichy regime were purged in a two-stage process where each case was judged twice by two different courts. Using a difference-in-differences strategy, we show that Law graduates, a powerful social group in French politics with strong connections to one of the two courts, had a clearance rate that was 10 percentage points higher than others. This facilitated the persistence of that elite group. A systematic analysis of 17,589 documents from the defendants' dossiers is consistent with the hypothesis that the connections of Law graduates to one of the two courts were a major driver of their ability to avoid the purge. We consider and rule out alternative mechanisms.

Reference Details

CWPE 2232
Published 18 May 2022

Key Words Purges, Political transitions, Elite persistence, Connections
JEL Codes D73, K40, N44, P48

Website www.econ.cam.ac.uk/cwpe

The Origins of Elite Persistence: Evidence from Political Purges in post-World War II France*

Toke S. Aidt†, Jean Lacroix‡, Pierre-Guillaume Méon§

Abstract

This paper studies a new mechanism that allows political elites from a non-democratic regime to survive a democratic transition: connections. We document this mechanism in the transition from the Vichy regime to democracy in post-World War II France. The parliamentarians who had supported the Vichy regime were purged in a two-stage process where each case was judged twice by two different courts. Using a difference-in-differences strategy, we show that Law graduates, a powerful social group in French politics with strong connections to one of the two courts, had a clearance rate that was 10 percentage points higher than others. This facilitated the persistence of that elite group. A systematic analysis of 17,589 documents from the defendants' dossiers is consistent with the hypothesis that the connections of Law graduates to one of the two courts were a major driver of their ability to avoid the purge. We consider and rule out alternative mechanisms.

Keywords: Purges, Political transitions, Elite persistence, Connections
JEL Codes: D73, K40, N44, P48

*This project receives financial support from the British Academy/Leverhulme Small Research Grants SRG 2018. Olivier Wieviorka contributed greatly at the project's early stages by sharing data, by offering guidance for our archival work and by discussing the functioning of the *Jury d'Honneur*. We thank Julia Cagé, Tommaso Giommoni, Marc Goni, Emiliano Grossman, Tommy Krieger, Volker Lindenthal, Andreas Madestam, Jean-Laurent Rosenthal, Argyris Sakalis, Hans-Joachim Voth and Vincent Pons for comments and suggestions. We also benefited from comments by seminar participants at the Paris School of Economics, at the Université Paris-Saclay, at the London School of Economics, at the Université Paris Panthéon-Sorbonne, at the 2022 ASSA meetings (EHA Session), at the 2021 Economic History Association Meeting (EHA - Tucson), at the workshop on the "Political Economy of Wars and political transitions" (Cambridge), at the 2021 SIOE Conference, at the 2021 Spring Meeting of Young Economists (Bologna), at the 2021 French Economic Association Meetings (AFSE - Lille), at the 2021 European Public Choice Society Meetings (EPCS - Lille), at the 14th CESifo Workshop in Political Economy, at the 2021 ZEW Workshop in Political Economy, at the 2021 Political Economy of Democracy and Dictatorship workshop, at the 2021 Silvaplana Workshop and at the EPAB workshop. We would like to thank Camille Besse, Louise Dutertre and Victor Westergren for their wonderful work as research assistants. All remaining errors are ours.

†tsa23@econ.cam.ac.uk, Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, United Kingdom and CESifo Munich.

‡jean.lacroix@universite-paris-saclay.fr, Université Paris-Saclay, Faculté Jean Monnet, RITM, 54 boulevard Desgranges, 92330 Sceaux

§pgmeon@ulb.ac.be, Centre Emile Bernheim, Université libre de Bruxelles, Av. F.D., Roosevelt, 50, 1050, Brussels.

1 Introduction

Political elites persist and are often able to remain influential following transitions from autocracy to democracy. For instance, the transition to democracy in Eastern Europe in the 1990s did not eliminate the political power of the old communist elite (Pakulski et al., 1996; Linz et al., 1996). In Indonesia, many of the mayors who had already served under the Soeharto regime were elected again after the transition to democracy and even stayed in office longer than other mayors (Martinez-Bravo et al., 2017). In post-Pinochet Chile, González et al. (2021) also observe that mayors appointed by Pinochet held an electoral advantage in municipal elections after the 1990 democratic transition. Historically, the sequence of franchise extensions in the United Kingdom between 1832 and 1885 did little to break the British aristocracy's monopoly on power (Berlinski et al., 2014). Likewise, after the US civil war and the enfranchisement of African Americans, the Southern white elite managed to maintain *de facto* power for many years (Acemoglu and Robinson, 2008; Besley et al., 2010). How can elites often compromised by their association with previous autocratic regimes continue to dominate politics after a transition to democracy? The answer to this question is important because elite persistence creates political inequality (Van Coppenolle, 2020) and serves as a roadblock to policy reform (Acemoglu and Robinson, 2006).

One possible answer is that elite groups have a comparative advantage over larger groups due to their small size and because of concentrated benefits (Olson, 1965). This allows them to invest in *de facto* power that enable them to survive the loss of *de jure* power associated with a transition to democracy (Acemoglu and Robinson, 2006, 2008). While elite groups have an incentive to invest in maintaining their political power, how they do it is less clear. Acemoglu and Robinson (2008) conjecture that they use wealth or weapons. Michels (1968) emphasized the control of communication technologies and political skills. Besley et al. (2010) focus on the elimination of political competition. In this paper, we document an alternative mechanism: connections. We argue that members of a former, compromised, elite can leverage connections to members of the new, uncompromised, elite to survive the transition from autocracy to democracy. By connections we mean both shared social ties and the use of these social ties. Those connections, built before the transition among alumni, colleagues, friends, or relatives, can provide support when the new regime decides what to do with the compromised elite. In short, these connections determine who is purged and who is not.

We document the role of such connections in the purge that took place after World War II in France.[1] The liberation of France meant an abrupt transition from the Vichy regime, a dictatorship that had cooperated with Nazi Germany, back to a democratic republic (the Fourth Republic). To reinstate democracy, the post-war authorities had to purge politicians who had collaborated with the Vichy regime and determine who would be allowed to continue a political career. Three features of the purge allow us to systematically investigate the role of connections in elite persistence. The first feature is that there was a well-identified group of politicians to purge. It included the parliamentarians who had endorsed, in a vote on July 10, 1940, the enabling act that cleared the way for the Vichy regime. Yet, some of these parliamentarians later took active part in the resistance, sometimes shortly after the vote. The new post-war authorities, therefore, had to sift away real supporters of the Vichy regime from those who had given in to pressure at the time of the vote but had later redeemed themselves. The second feature of the purge is that it followed a structured legal process. Specifically, two bodies were tasked with reviewing the cases sequentially. A case was first reviewed

[1] Technically, this is a case of "lustration". Lustration is a procedure whereby a legal body examines the actions of individuals in order to prevent those compromised with a previous regime to hold a category of positions in a new regime (see, e.g., Kaminski and Nalepa (2006), Nalepa (2008) and Bates et al. (2020)). After the war, France adopted an "accusation-based truth-revelation" lustration procedure that resembled usual court proceedings. For simplicity, we refer to this as a"purge" rather than as lustration.

by a local *Comité départemental de libération* (hereafter the CDLs), established in each French department with members recruited from local resistance groups. In the second stage, each case was reviewed by the *Jury d'Honneur* (hereafter the *Jury*) in Paris, consisting of three prominent members of the resistance with a Law background. The *Jury* could either follow the judgment of the CDLs or overrule it. This two-stage process meant that each case was heard twice. Importantly, the dossiers on each defendant in the archives of the *Jury* allow us to document connections between each defendant and his supporters and the *Jury*. The third feature of the French post-war purge is that there was a well-identified elite group in French politics whose connections to the *Jury* we can study. Specifically, both before and after the war, many French politicians were Law graduates (Le Béguec, 2003). Law graduates formed a tightly knit social group, with connections first established during their studies and subsequent training, and later in their careers, maintained through the Bar Association, clubs, and speech contests. We conjecture that those connections were instrumental in interceding with the three members of the *Jury*, who all had connections to Parisian faculties of Law.

Based on those three features – a target group to purge, a well-defined elite with connections to the *Jury*, and a two-stage legal process – we adopt the method developed by Anwar and Fang (2006) and Alesina and La Ferrara (2014) to study the racial bias of US courts to identify the advantage of Law graduates in front of the *Jury* using the decisions of the CDLs as a counterfactual. The difference-in-differences estimate of the acquittal rates across the two courts provides a measure of the relative *Law graduate advantage* in front of one court compared to the other. The main result is that the difference in acquittal rates between Law graduates and other defendants was 10 percentage points higher in front of the *Jury* than in front of the CDLs. We argue that this Law graduate advantage before the *Jury* was due to the connections among Law graduates. From primary archival research, we created an inventory of the 17,589 documents contained in the dossiers of the defendants in the *Jury*'s archives. We find that the dossiers of Law graduates contained more letters of support from establishment figures who were well-connected to the *Jury,* that many of them were written in an informal style that suggests a personal connection to the receiver of the letter, and that the Law graduate advantage disappears when we control for measures of connections in the difference-in-differences regression. This is consistent with the claim that the Law graduate advantage in front of the *Jury* emerged because they benefited from connections that could be leveraged to lobby on their behalf.

Our analysis speaks to four strands of literature. First, we contribute to the literature on elite persistence after democratic transitions (O'Donnell and Schmitter, 2013). We document a new mechanism – connections – explaining why elites persist and survive democratic transitions that complements existing explanations (e.g., Higley and Burton, 1989; Acemoglu and Robinson, 2006; Martinez-Bravo, 2014; Martinez-Bravo et al., 2017; González et al., 2021).[2] Second, we shed light on an understudied type of political purges: purges during democratic transitions. The focus of the existing literature is on purges in autocracies where purges or the threat thereof can protect an autocrat from coups originating from within the ruling coalition (Svolik, 2009; Bueno de Mesquita and Smith, 2017; Montagnes and Wolton, 2019; Goldring and Matthews, 2021). While political purges after a democratization also aim at consolidation, they are fundamentally different from those in autocracies because they are constrained by the rule of law and must happen within a legal framework. We provide a theoretical conceptualization of this process and new empirical evidence on how

[2]The sources of elite persistence within *given* institutional structures are reasonably well-established. The classical argument of Michels (1968) is that any complex social organization will eventually be dominated by a small elite because leaders control resources that rank-and-file members do not: superior information, communication technologies, and political skills. In most Western democracies, incumbency advantage (Eggers et al., 2015) and internal legislative procedures (Berlinski et al., 2007) enables the same (type of) politicians to stay in power and these advantages can be passed on to family members (Querubin, 2015; Van Coppenolle, 2017; Fiva and Smith, 2017; Dal Bó et al., 2009). In autocracies, dynastic transitions develop as a norm to avoid the successor problem (Tullock, 1987; Kurrild-Klitgaard, 2000).

political purges can facilitate elite persistence that extend previous research (Ang and Nalepa, 2019). Third, we contribute to the literature on the effect of connections in economics and politics (e.g., Fisman, 2001; Dal Bó and Di Tella, 2003; Cohen and Malloy, 2014; Wolton, 2017). We show that connections can help elite groups surviving transitions from autocracy to democracy. Fourth, we contribute to the literature on bias in sentencing (Voeten, 2008; Shayo and Zussman, 2011; Alesina and La Ferrara, 2014; Lim et al., 2015; Park, 2017). We document how connections of defendants to a court can result in more leniency.

The rest of the paper is structured as follows. Section 2 presents the historical background of the purge. Section 3 presents a theoretical framework that illustrates how connections can influence court decisions and allows us to develop our identification strategy. Section 4 describes the main dataset and reports the baseline results documenting the Law graduate advantage. Section 5 presents evidence that the source of the Law graduate advantage is connections by drawing on the information from dossiers of the defendants. Section 6 concludes.

2 Historical background: Political purges in post-World War II France

This section, firstly, describes the historical facts related to the transition from the Vichy regime to the Fourth Republic and to the political purge that took place during this transition. Secondly, it documents the prominent role that Law graduates played in French politics both before and after World War II and that they formed a well-connected social group. Finally, based on the dossiers of the defendants from the archives of the *Jury*, it presents anecdotal evidence of the role connections played in framing the cases put to the *Jury*.

2.1 The transition from the Vichy Regime to the Fourth Republic

On 10 July 1940, in the wake of the French military defeat in the Battle of France, the French Parliament passed an enabling act granting full power to Marshall Philippe Pétain. Until the liberation of France by the Allies, the Vichy regime was nominally in charge of the civil administration of the country, even though the country was first partly then fully occupied. The regime was a dictatorship. It implemented a radical anti-modern reform program and sided with Germany and Italy and collaborated with them in their fight against the resistance and in persecuting Jews (Paxton, 1972). The regime collapsed as the allied troops liberated France and the Vichy government was eventually forcibly moved to Germany to serve as a puppet government. While the Vichy regime was nominally ruling over mainland France, a provisional government had emerged from the various branches of Free France led by General Charles De Gaulle and the French colonies (Paxton, 1972; Albertelli et al., 2019) and, at the end of the war, the Provisional Government of the French Republic ("Gouvernement provisoire de la République française" or the "GPRF") was established. The GPRF was to tasked with dismantling the Vichy regime and restoring a democratic republic.

2.2 The purge

As large parts of French society had been compromised by the Vichy regime, a purge was needed at all levels, from civil servants and politicians to writers, journalists, and intellectuals.[3] On 21 April 1944, the GPRF published an order rendering ineligible for election to public office various groups of individuals who had compromised themselves. The order explicitly singled out parliamentarians who had voted for the enabling act. (Paxton, 1972).

A total of 669 parliamentarians, both deputies and senators, had taken part in the vote on the enabling act. Out of them, 80 voted against the act, 20 abstained, and 569 voted for (Lacroix et al., 2023). By default, the order of 21 April 1944 banned the latter from participating in post-war politics. However, the ban could be waived if a parliamentarian could prove that he had taken active part in the resistance (Wieviorka, 2001). Until 6 April 1945, departmental prefects could waive the ban and did so for 51 parliamentarians for whom there was indisputable evidence of participation in the resistance (Wieviorka, 2001).

An order of 6 April 1945 describes in detail the procedure to be followed to waive the ban for all the remaining (banned) parliamentarians (hereafter the parliamentarians). Each case underwent a two-stage legal process and was assessed sequentially, first, by a local and, second, by a national court. At the first stage, the case was considered by "Comités départementaux de libération" which operated in each department, France's main administrative unit. The CDLs had been created by the resistance. Their composition is unknown but likely reflected the balance of power of local resistance groups and the CDLs were, therefore, heterogeneous in their makeup (Albertelli et al., 2019). One of their tasks was to assess the dossiers filed by parliamentarians who wanted their ban from politics waived.

At the second stage, each case was assessed by the *Jury d'Honneur*, a national court established specifically to purge former supporters of the Vichy regime from politics, including all parliamentarians who had voted in favor of the enabling act. It could overrule the decisions of the CDLs. The *Jury* had three members. René Cassin, vice-President of the *Conseil d'Etat*, presided over the Court. He was assisted by Maxime Blocq-Mascart, representing the *Conseil National de la Résistance,* and by André Postel-Vinay, representing the *Ordre de la Libération*. The *Jury* used one criterion to acquit the parliamentarians: "an active (and direct) participation in resistance activities before November 1942".[4] Although none of the orders establishing the *Jury* defined how decisions were to be taken, they were collegiate and announced as a consensus decision.

All the members of the *Jury* were Law graduates. Its chairman, René Cassin, was a lawyer at the Paris bar and a Law professor in Paris. He was a prominent figure in the legal milieu during and after the war. The *Dictionnaire historique de la Résistance* (Marcot, 2006) refers to him as "the jurist of Free France" (p. 383). André Postel-Vinay held a bachelor degree in Law and had studied at "Ecole libre des sciences politiques (Sciences-Po)" in Paris, where students study a blend of social sciences. The school, created in 1871 by a Professor of Law, Emile Boutmy, was, however, oriented towards the study of Law and had close connections to Law Faculties, as evidenced historically by the composition of its Faculty.[5] Maxime Blocq-Mascart was a graduate of the same school. In addition, the *Jury* was assisted by rapporteurs assigned to each case and also by administrative staff. That staff mainly came from the *Conseil d'Etat*, the highest administrative court in

[3] For more historical information on the purges, see Wieviorka (2001); Baruch (2003); Elster (2006).

[4] Minutes of the first meeting of the *Jury* (quoted in Wieviorka 2001, chap. 5).

[5] Emile Boutmy, for example, published a study on constitutional laws (https://gallica.bnf.fr/ark:/12148/bpt6k235741/f1.item.texteImage - Consulted February 25, 2021). A book celebrating the 25 years of the creation of the school (in 1896) shows that both the President of the Board (M. Léon Aucoc) and the General Secretary of Faculty members (M. C. Dupuis) were Law graduates (Source: https://gallica.bnf.fr/ark:/12148/bpt6k96193204/f9.item - Consulted February 25, 2021).

France, and the *Jury* was located in the building of the *Conseil d'Etat*. In short, the *Jury's* members and its staff had a Law background and strong connections to one of the most influential groups in French politics at the time, Law graduates.

2.3 "The Lawyers' Republic"

Law graduates formed an influential social group for two reasons. First, its structure, with the Bar association, its clubs and speech contests, meant that Law graduates bond during their training and later on cultivate these connections within the Bar association. Second, Law graduates were historically tightly-linked to French politics. For example, many lawyers studied Law in combination with political science. Specifically, a substantial share of parliamentarians in the National Assembly were lawyers. They represented 19.6 percent of parliamentarians in the 1936-1940 National Assembly. Furthermore, they often held positions of power. From January 1920 to March 1940, France had 19 Council Presidents.[6] Among them, 13 were lawyers (Le Béguec, 2003). During the French Third Republic, Law graduates, hence, formed what Le Béguec (2003) calls the "Lawyers' Republic". After the War, the proportion of lawyers in the Assembly decreased slightly but still amounted to 15.6 percent in 1958 (Le Béguec, 2003). In addition, their influence remained substantial. From 1946 to 1958, two of the four Presidents of the National Assembly had a Law degree. In addition, the first President of the Council of the Republic, the Upper Chamber under the Fourth French Republic, Auguste Champetier de Ribes, was a Law graduate. In short, in 1945, Law graduates were a powerful and well-defined social group that assumed a prominent role in French politics. Although the proportion of lawyers in the Assembly decreased after the war, many survived the post-war purge and secured influential positions. We conjecture that the capacity of Law graduates to survive the regime change was related to their proximity to the *Jury*, whose members, as noted, had a Law background.

2.4 Connections and the post-war purge: Archival evidence

The Archives Nationales de France hold the dossiers of the defendants who were judged by the *Jury*. They provide a wealth of contextual evidence that connections, not only were considered by many defendants to have biased the *Jury,* but also that connections were frequently leveraged to lobby its members.

Firstly, a common line of defense used by defendants to refute the decision of the *Jury* to purge them from politics was that undue pressure had biased the *Jury's* decisions. For example, in their statement of defense, defendant A and B directly accused the *Jury* of being partial and the whole purge process to be biased.[7] Some other defendants also tried to bypass the decision of the *Jury* by complaining about its partial nature. On 18 October 1945, defendant C wrote in a letter to General De Gaulle: "It is not about justice, but about connections".[8] Conversely, some supporters of defendants also used the rhetoric of partial decisions to get them acquitted. For example, defendant's D dossier contains an anonymous letter denouncing "a political plot against" him.[9]

[6] This is equivalent to the position of Prime Minister.

[7] Names are anonimized as archives are still-classified and we refer to defendants by letters. Archives Nationales de France, Reference: AL//5308 and AL//5309.

[8] Archives Nationales de France, Reference: AL//5324.

[9] Archives Nationales de France, Reference: AL//5321.

Secondly, many individual dossiers have direct evidence of how connections were leveraged. A perfect illustration is the letter defendant E sent to René Cassin on 6 June 1945 to ask for a reappraisal of his case. He writes "I took the liberty to ask you this favor, because many have advised me to use my contacts".[10] In the dossier of defendant F, there is a letter from one of his supporters asking Fedia Cassin, brother of the president of the *Jury,* if he could help him reach out to René Cassin.[11] The dossiers also show how defendants mobilized their own connections to contact the *Jury* to lobby on their behalf. For example, the leader of a resistance network wrote to defendant G: "I would like to let you know that after learning about the injustice concerning your case, I personally went to see M. Bernard (Rapporteur on the case)."[12] Similarly, the dossier of defendant H contains a note from the cabinet of General De Gaulle forwarding a letter from the defendant to the Ministry of the Interior. The note states "It looks like the case of defendant H deserves some more attention" as the defendant has previously been purged from politics.[13] Finally, in some cases, the dossiers also show defendants contacting their connections to seek help. In this vein, defendant I, in a letter to the president of his political group, wondered if he should go to the *Jury* with an introductory note from him.[14]

Thirdly, we observe various degrees of informality in the way the letters in the dossiers address the recipient, suggesting various degrees of familiarity with the members of the *Jury.* For instance, several dossiers include letters from influential figures addressed to René Cassin and with salutations such as "Mon cher ami" ("My dear friend"). For example, the dossier of defendant J includes a letter which a supporter wrote on 31 July 1945 to René Cassin with that salutation stating that not acquitting defendant J would be a mistake.[15] Similarly, the dossier of defendant K contains a letter from the defendant to the President of the Constituent Assembly on April 1946 asking for support. It is followed by a letter from the President of the Constituent Assembly to René Cassin on October 1946 starting with "Mon cher ami" and asking for a new assessment of the case of defendant K.[16]

In conclusion, we observe clear evidence from the dossiers of the defendants that connections played a key role in the interactions with the *Jury.* It remains to determine whether and how such connections systematically influenced the decisions of the *Jury* and, more specifically, how they operated for different groups of defendants. That is the task of the rest of the paper. We start with a new theory of court decisions and connections from which we develop our empirical strategy before we turn to the statistical evidence.

3 Theory

3.1 Model overview

The model portraits a situation where defendants who voted for the enabling act are considered sequentially by a lower (the CDLs) and an upper court (the *Jury*). The courts have to decide if the ban on political participation already imposed on the defendants stands or is overturned. Each court defines its standard of proof required for acquittal. Thereafter, each court evaluates the evidence and the defendants or their

[10] Archives Nationales de France, Reference: AL//5298.
[11] Archives Nationales de France, Reference: AL//5298.
[12] Archives Nationales de France, Reference: AL//5298.
[13] Archives Nationales de France, Reference: AL//5331.
[14] Archives Nationales de France, Reference: AL//5334.
[15] Archives Nationales de France, Reference: AL//5303.
[16] Archives Nationales de France, Reference: AL//5311.

supporters can report mitigating circumstances with the aim of influencing the court decisions. This is the novel aspect of the model. The communication of mitigating circumstances is cheap talk, so there is no guarantee that it will, in fact, influence the courts. The fundamental problem is that the defendants and their supporters have an incentive to claim mitigating circumstances whether there are such circumstances or not. We model two mechanisms that can overcome this problem. Both mechanisms are related to connections between the courts, on the one hand, and the defendants and their supporters, on the other. The first mechanism is *direct* connections between the defendant and the courts. The second is *indirect* connections operating via a third party (a supporter of the defendant). We show that both of these mechanisms can under certain circumstances help defendants with connections get acquitted by the court to which they are connected.

3.2 Setting the bar

Two courts, indexed by $c \in \{L, U\}$, are tasked with judging a fixed number of defendants indexed by $i \in D$. The lower court L (the CDLs) hears each case first and its judgment can be confirmed or overturned by the upper court U (the *Jury*). Before any evidence is heard, the two courts set their own bar for acquittal. After that, evidence is presented to the courts, they receive letters from the defendants and/or their supporters claiming mitigating circumstances and make their decisions.

Let x summarize the evidence presented to a court with $x \in (-\infty, \infty)$. A defendant who presents evidence stronger than the bar set by that court will be acquitted. To set the bar, we assume, as in Alesina and La Ferrara (2014), that the court's aim is to avoid making type 1 (convicting innocent defendants) and type 2 (not convicting guilty defendants) errors.[17] The weights that court c puts on type 1 and 2 errors are α_c and $1 - \alpha_c$, respectively. The evidence presented to the courts is drawn from the cumulative distribution function $A_G(x)$ if the defendant is guilty (did not participate in the resistance) and from $A_I(x)$ if innocent (participated in the resistance) and the corresponding density functions are a_G and a_I. We assume that the defendants can be divided into sub-groups based on fixed observable characteristics, such as profession, political affiliation, region of residence, age, religion, and other observable characteristics, and that the proportion of guilty defendants in those subgroups may be perceived by the two courts to be different. We let $g \in \{1, 2, ...N\}$ with $\cup_g D_g = D$ index these sub-groups. The two courts assume that the proportion of guilty among defendants belonging to sub-group g is π_g. The objective function of court c can, then, be written as

$$\min_{x(c,g)} \sum_{g=1}^{N} \alpha_c(1 - \pi_g)A_I(x) + (1 - \alpha_c)\pi_g(1 - A_G(x)). \tag{1}$$

The optimal bar for court c for defendants belonging to sub-group g is the solution to

$$\frac{\alpha_c}{1 - \alpha_c}\frac{1 - \pi_g}{\pi_g} = \frac{a_G(x^*(c, g))}{a_I(x^*(c, g))} \tag{2}$$

and denoted $x^*(c, g)$. Given the bar, the probability that a defendant belonging to group D_g is being acquitted

[17] This is consistent with the type of democratic purge that we study. A purge in an authoritarian regime would likely give little weight to type 1 errors, if any.

before court c is

$$Pr[c,g] = \pi_g(1 - A_G(x^*(c,g)) + (1-\pi_g)(1 - A_I(x^*(c,g)) \tag{3}$$

$$\equiv 1 - H_g(x^*(c,g)), \tag{4}$$

where $H_g = \pi_g A_G + (1-\pi_g)A_I$. After applying a first order linear approximation, we can write the probability of acquittal as

$$Pr[c,g] \approx a + b_c + b_g. \tag{5}$$

We observe that a defendant's chances of acquittal differ before the two courts for two reasons. First, the courts may weigh the risk of the two types of mistakes differently (differences in α_c). This would lead to systematic differences in the acquittal rates between the courts and is captured by b_c in the linear approximation. This represents differences in the objectives and procedures of the two courts that can lead to differences in acquittal rates. Second, the two courts may perceive, based on observable characteristics, some groups of defendants to be more likely to be guilty than others (due to differences in π_g). This is a manifestation of statistical discrimination and can explain systematic differences in acquittal rates between different sub-groups of defendants and is captured by b_g in equation (5). The next section models how the defendants and their supporters can use connections to claim mitigating circumstances to increase the chance of acquittal. We show how this can lead to systematic differences in the acquittal rates for connected and unconnected defendants.

3.3 Connections

After the bars have been set, the defendants or their supporters can present evidence of mitigating circumstances to the courts. For each defendant i, we assume that there may or may not be mitigating circumstances. We denote this by θ which can take two values: θ_Y if there are mitigating circumstances and θ_N if not with $\theta_Y > \theta_N$. This is private information to the defendants and their supporters and not known to the courts and cannot be externally verified. The defendants or their supporters may present evidence of mitigating circumstances to the courts by sending letters. We denote the content of a letter sent by defendant i either personally or via a supporter to court c by $l_{i,c} \in \{\theta_Y, \theta_N\}$. If the court accepts a letter claiming that $\theta = \theta_Y$ for a defendant i, then it updates its belief about the cumulative distribution function from which the evidence for that defendant is drawn from $H_g(x)$ to the cumulative distribution function $\bar{H}_g(x)$. We assume that $H_g(x)$ is first-order stochastically dominated by $\bar{H}_g(x)$, i.e., $H_g(x) \leq \bar{H}_g(x)$ for all x with strict inequality over some interval. For simplicity, we write:

$$\bar{H}_g(x^*_{c,g}) = H_g(x^*_{c,g}) - \eta \tag{6}$$

with $\eta > 0$. In other words, the mitigating circumstances make it more likely that defendant i is acquitted and the effect is increasing in η. We can interpret a letter $l_{i,c} = \theta_N$ as not sending a letter containing arguments about mitigating circumstances to court c. The objective of court c is to base its judgment on all the facts and it needs to decide if mitigating circumstances should be taken into account or not. Formally, the objective of court c in relation to the case of defendant i is to minimize mistakenly taking mitigating circumstances into account: $U_{i,c} = -(m_{i,c} - \theta)^2$ where m_i is interpreted as a decision to take mitigating circumstances

into account ($m_{i,c} = \theta_Y$) or not ($m_{i,c} = \theta_N$) for defendant i. The optimal decision is $m_{i,c} = E(\theta|l_{i,c})$ where E is the expectation operator. If the court gets no letter for a defendant i, then its prior is that there are no mitigating circumstances and $m_{i,c} = \theta_N$. All the defendants want mitigating circumstances, if any, to be taken into account by the courts, but also to be acquitted. The objective of defendant i with information θ, therefore, is $U_{i,c} = -(m_{i,c} - \theta - \gamma)^2$, where $\gamma > 0$ captures the desire to be acquitted, i.e., the optimal decision by court c from the point of view of defendant i, $m_{i,c} = \theta + \gamma$, is biased in favor of taking mitigating circumstances into account whether there are, in fact, such circumstances. We model the interaction between a defendant i (or a supporter of defendant i, respectively) and court c as a sequential game of asymmetric information where the defendant, firstly, sends a letter to the court which, secondly, updates its beliefs about whether there are mitigating circumstances based on the content using Bayes rule where possible. The equilibrium concept is perfect Bayesian equilibrium. We assume that the motive to falsely claim mitigating circumstances is so strong that it is not possible, in general, for any defendant to write a letter that credibly claims mitigating circumstances. Formally, we assume

$$\gamma > \frac{\theta_Y - \theta_N}{2} \equiv \bar{\gamma} \tag{7}$$

The difference $\theta_Y - \theta_N$ can be interpreted as a measure of how much the defendants care about a fair trial. Accordingly, assumption (7) says that all defendants care so much about being acquitted relative to a fair trial that they cannot communicate mitigating circumstances directly to the court. Formally, the assumption rules out informative perfect Bayesian equilibria of the type considered by Crawford and Sobel (1982) in which the defendants write a letter claiming mitigating circumstance only when there are, in fact, such circumstances. The court, therefore, sticks to its prior: there are no mitigating circumstances. In order to influence the court, a defendant needs to use his direct or indirect connections to the courts.

3.3.1 Direct connections to the courts

All defendants can communicate with the courts but a defendant with a *direct* connection to a court has an advantage in doing so. To model direct connections, we assume that communication is associated with a fixed cost.[18] A direct connection lowers the cost of communication. There can be many reasons for this, ranging from personal or professional ties to shared knowledge about the proper etiquette for communication in the context. Each defendant i got a connection to court c indexed by a fixed cost of using it, $f_{i,c}$. The first proposition shows when and how direct connections can help a defendant get acquitted before court c.

Proposition 1. *There exist two values $\bar{f}$ and $\underline{f}$ with $\bar{f} > \underline{f}$ such that*

1. **Ineffective connected defendants**: Defendants with a cost of communicating $f_{i,c} < \underline{f}$ cannot convince court c and will not submit a letter claiming mitigating circumstances to that court.
2. **Effective connected defendants**: Defendants with a cost of communicating $\underline{f} \leq f_{i,c} \leq \bar{f}$ will submit a letter claiming mitigating circumstances ($l_{i,c} = \theta_Y$) to court c if and only if that is the case and the court will believe the claim.
3. **Unconnected defendants**: Defendants with a cost of communicating $f_{i,c} > \bar{f}$ will not submit a letter claiming mitigating circumstances to court c even through the court would believe such a letter.

[18] See Grossman and Helpman (2002) for a similar approach to lobbying.

Proof. See Appendix A.1 □

The proposition shows that the defendants are endogenously sorted into two subsets: the set of connected defendants, denoted K_c^d, and the set of unconnected defendants, denoted k_c^d, where the superscript d refers to direct connections. The set K_c^d of connected defendants consists of two types of defendants. While all connected defendants have relatively low cost of communicating with the court ($f_{i,c} \leq \bar{f}$), only some of them are successful at convincing court c that mitigating circumstances are relevant. If they are "too" connected ($f_{i,c} < \underline{f}$), then the court will not trust their letters. Indeed, in these cases, the cost of sending information is so low that defendants will do it regardless of whether they are innocent or not. As a consequence, receiving a letter from them is uninformative. The defendants who can influence the court are those with a moderately good connection ($f_{i,c} \in [\underline{f}, \bar{f}]$). The court will believe what they have to say and so this group will submit letters claiming mitigating circumstances. For the set of unconnected defendants the cost of communication is too high ($f_{i,c} > \bar{f}$) and they do not submit letters claiming mitigating circumstances to the court. The proposition has the following empirical implication: if the population of defendants contains directly connected defendants as well as unconnected ones, then, on average, the connected defendants are at least as likely to be acquitted as the unconnected and strictly more likely if they are not all "too" connected.

3.3.2 Indirect connections

A defendant may have an indirect connection to the courts through a third party who is himself connected to them. The third party observes θ and thus knows whether there are mitigating circumstances or not. We assume that the third party connected to defendant i has the following objective function

$$U_{i,c}^T = \beta_{i,c} U_{i,c} + (1 - \beta_{i,c}) U_i = -\beta_{i,c}(m_{i,c} - \theta)^2 - (1 - \beta_{i,c})(m_{i,c} - \theta - \gamma)^2. \tag{8}$$

The assumption is that third party (T) linked to defendant i partly aligns with the objective that court c ($U_{i,c}$) uses to judge mitigating circumstances and partly with the objective of the defendant (U_i), with the weight $\beta_{i,c}$ determining the relative weight on the two which may vary for defendants across the two courts. The optimal decision of court c in case i from the point of view of the third party is $d_{i,c}^T = \theta + (1 - \beta_{i,c})\gamma$. The third party is, by definition, connected to the courts and can thus send a letter at low cost (for simplicity, we set the cost at zero for both courts, i.e., $f_c^T = 0$ for all c). One interpretation of this is that some third parties intrinsically care about the integrity of the legal process. Another is that the judges may trust people they know better more – in part because these people would lose more (in terms of reputation) if the evidence they provided proved to be wrong.

Proposition 2. *There exists a $\bar{\beta} \in (0, 1)$ such that a third party with $\beta_{i,c} > \bar{\beta}$ who submits a letter on behalf of defendant i can convince court c that there are mitigating circumstances for defendant i when that is the case. A letter from a third party with $\beta_{i,c} \leq \bar{\beta}$ is not believed by court c.*

Proof. See Appendix A.2 □

The proposition says that third parties are effective at intervening on behalf of a defendant before court c if they at least to some degree share the same objective as the court. This splits the set of defendants into two subsets: those with indirect connections K_c^{id} and those without k_c^{id}, where superscript id refers to

indirect connections. A letter from a third party who mostly cares about the defendant will not, in general, influence a court's decision. The advantage of having a third party intervening on behalf of a defendant is that third parties (to varying degrees) are concerned about the integrity of the legal process and thus have objectives that partly overlap with those of the court. This is what makes their letters of support credible and influential. The empirical predictions that flow from this is that defendants who are supported by connected third parties are more likely to be acquitted than those who are not, and that these defendants would have more letters of support in their case files from third parties with a connection to the court.

From theory to empirical strategy

To test the predictions and to estimate the causal effect of connections, we exploit that we have information on the judgment of the two courts for each defendant. This enables us to use a difference-in-differences strategy to isolate the relative effect of connections on the probability of acquittal. To explain the logic of the test, let us focus on direct connections. The logic is the same for indirect connections. Our aim is to estimate η in equation (6). Doing so requires that the "connectedness" of the defendants varies by court so that we can separate the effect of connectedness from other group characteristics that may induce statistical discrimination in the court decisions and from court-specific differences in acquittal rates related, for example, to the composition of the judges. To see how this works, suppose that no one has a direct connection to the lower court (L), i.e., $f_{i,L} > \bar{f}$ for all i. Using equation (5), the probability of acquittal is $a + b_L + b_g$ for all $i \in D_g$ and all g. In contrast, assume for the upper court (U) that the cost of communication is such that the defendants, according to proposition 1, are endogenous sorted into the set of connected defendants K_U^d and the set of unconnected defendants k_U^d. Finally, to illustrate how we can eliminate the effect of statistical discrimination against or in favor of connected defendants, assume that there is a fixed characteristic observed by the two courts that correlates perfectly with being connected to the upper court. That is, we divide the defendants into two sets ($N = 2$) based on this feature such that $D_1 = K_U^d$ and $D_2 = k_U^d$. In the empirical application this characteristics is being a Law graduate. The probability of acquittal at the upper court, then, is $a + b_U + b_2$ for $i \in k_U^d = D_2$ (the unconnected) and $a + b_U + b_1 + I_i\eta$ for $i \in K_U^d = D_1$ where I_i is an indicator function equal to 1 if defendant i is, in the sense of proposition 1, effectively connected ($\underline{f} \leq f_{i,U} \leq \bar{f}$) to the upper court and zero if the defendant is ineffectively connected ($f_{i,U} \leq \underline{f}$). Given these assumptions, the differences in the expected acquittal rates for connected and unconnected defendants in the two courts are

$$\Delta_L = E_{i \in K_U^d = D_1} Pr(L,1) - E_{i \in k_U^d = D_2} Pr(L,2) = b_1 - b_2 \tag{9}$$

$$\Delta_U = E_{i \in K_U^d = D_1} Pr(U,1) - E_{i \in k_U^d = D_2} Pr(U,2) = b_1 - b_2 + E_{i \in K_U^d = D_1} I_i\eta \tag{10}$$

The difference-in-differences estimate, then, is $\Delta = \Delta_U - \Delta_L = E_{i \in K_U^d = D_1} I_i\eta$. We notice, firstly, that if the exact same defendants were connected to both courts, then we would not be able to identify η. Identification requires differential connections to the two courts and the estimate of Δ should be interpreted as the relative effect of connections to the two courts. Secondly, if the defendants are "too" connected in the sense of proposition 1, then $\Delta = 0$, i.e., being connected has no effect on the probability of acquittal: the connections are ineffective. By netting out both court characteristics and statistical discrimination, the difference-in-

differences estimator allows us to identify connections specific to a court-defendant pair.

As we do not directly observe who is connected and who is not, we test the theory in two steps. First, in the next section, we establish that a particular group of defendants – Law graduates – who were connected to the *Jury* but not to the CDLs – experienced a higher acquittal rate before the *Jury* than before CDLs. In the following section 5, we use the detailed information derived from archival research on the content of the defendants' dossiers to relate this Law graduate advantage to direct and indirect connections between Law graduates and the *Jury*.

4 The Law graduate advantage before the *Jury*

To test the prediction related to the Law graduate advantage, we combine biographical information on the age, profession, education, etc. of the parliamentarians who voted on the enabling Act in 1940 (Wieviorka, 2001; Lacroix et al., 2023) with information on which of them were tried after the war and the decisions reached by the CDLs and the *Jury*, respectively.[19] The data set comprises 798 decisions on the cases of 399 individual defendants along with their personal characteristics. Table 1 cross tabulates the decisions of the *Jury* and the CDLs. The two agreed in 81.0% of the 399 cases. Our tests leverage the cases where the CDLs and the *Jury* disagreed (the entries in boldface). These were mostly cases where the *Jury* acquitted defendants against the judgment of CDLs (17.0% of the cases) but in a few cases the *Jury* banned defendants that the CDLs wanted to acquit (2.0% of the cases).

Table 1: Cross tabulation of the two courts' decisions

	CDL for acquittal	CDL for ban	Total
Jury for acquittal	32	**68**	100
Jury for ban	**8**	291	299
Total	40	359	399

We start our analysis with some descriptive evidence of the Law graduate advantage (Section 4.1) before we turn to causal identification and results (Sections 4.2 and 4.3).

4.1 Descriptive evidence

Figure 1(a) reports mean comparisons of the acquittal rate of Law graduates and other defendants before the two courts, respectively. Before the CDLs, the acquittal rate was 8,2% for Law graduates and 10,7% for other defendants. This difference is not statistically different, suggesting that the CDLs did not treat Law graduates differently from other defendants. Before the *Jury*, on the other hand, the acquittal rate of Law graduates was 30,9% compared to 22,8% for other defendants. This difference is statistically significant, suggesting that the *Jury* in contrast to the CDLs did treat Law graduates more leniently than other defendants. Using Equation (9) from Section 3, the difference-in-differences estimate of the relative Law graduate advantage is (30,9-22,8)-(8,2-10,7)=10,6 percentage points.

[19]The data are obtained from Wieviorka (2001) and consolidated with information from the still-classified individual dossiers of the defendants from the archives of the *Jury* (References AL//5295 to AL//5334). We sincerely thank Olivier Wieviorka for sharing his data with us.

Figure 1(b) reports separately for Law graduates and other defendants the percentage of cases for which the decisions of the CDLs and the *Jury* differed. The left-hand side panel displays the percentage of cases in which a CDL's decision to acquit a defendant was overturned by the *Jury* for Law graduates and other defendants, respectively. The percentages are similar: 1.8% for Law graduates and 2.1.% for other defendants. The right-hand side panel shows the percentages of cases in which a CDL's decision to ban a defendant was overturned by the *Jury* for the two groups of defendants. The difference is striking: the *Jury* acquitted defendants that a CDL wanted to ban in 24.5% of the cases for Law graduates but only in 14.2% of the cases for other defendants. Clearly, Law graduates tended to have their ban overruled by the *Jury*.

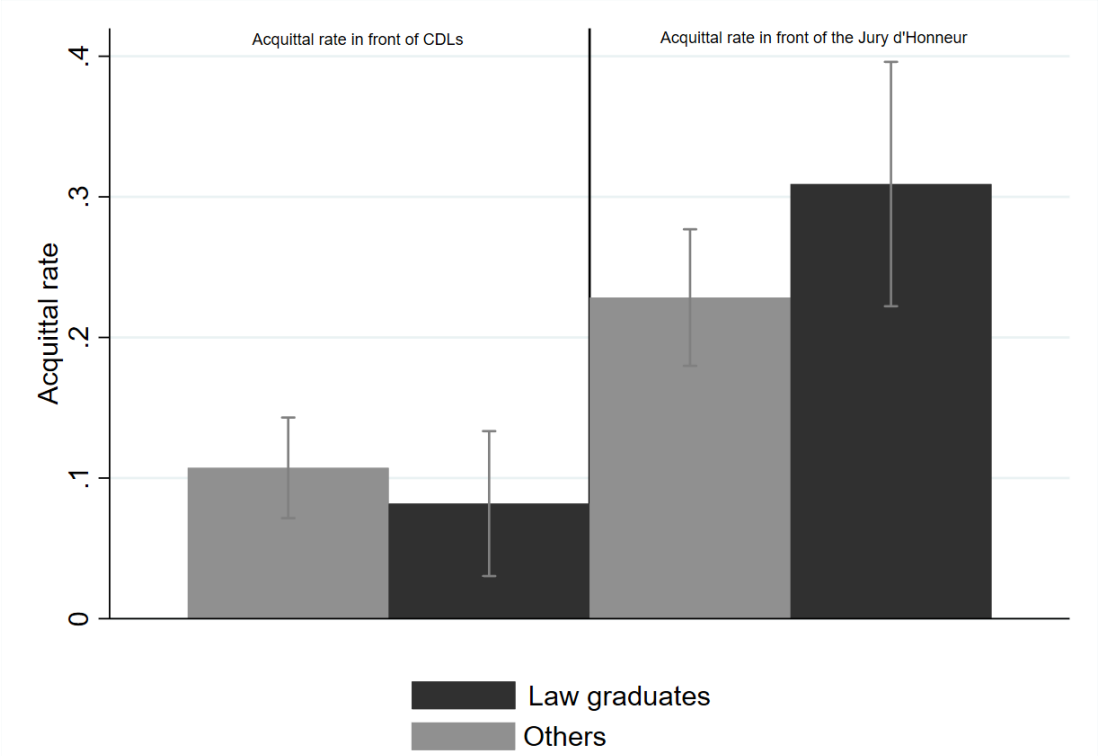
The difference-in-differences estimate of the *relative* Law graduate advantage could, in principle, reflect a disadvantage of Law graduates before the CDLs rather than an advantage before the *Jury*. However, two observations speak against this. First, from Figure 1(a), we see that the positive difference-in-differences estimate results from differential treatment of Law graduates before the *Jury* and not before the CDLs. Second, from Figure 1(b), we notice that the *Jury* overturns bans upheld by the CDLs more frequently for Law graduates than for other defendants; it does not overturn acquittal decisions by the CDLs more frequently for non-Law graduates than for Law graduates. In short, Law graduates were treated more leniently by the *Jury*. The relative Law graduate advantage can therefore, in practice, be interpreted as an absolute advantage of Law graduates before the *Jury*.

Another way to illustrate the Law graduate advantage is to ask if Law graduates were more likely than others to be acquitted by the *Jury* given the objective evidence related to their participation in the the resistance which was the official criterion for acquittal used by the *Jury*. To do this, we estimate, on the full sample of defendants, the probability that a defendant is acquitted by the *Jury* as a function of the information in his dossier regarding his participation in the resistance, controlling for the CDL judgment.[20] From this we can calculate the *counterfactual* acquittal rate for Law and non-Law graduates separately and compare them to the *actual* acquittal rates for the two groups. Figure 2 shows the result. The left-hand panel shows the counterfactual (in grey) and the actual acquittal rate (in black) for non-Law graduate defendants. Based on our counterfactual calculation, non-Law graduates had a 23.7% probability of being acquitted by the *Jury* compared with an actual probability of 22.8%. These are not statistically different: the *Jury* treated non-Law graduates exactly as we would expect given the evidence in their dossiers and the prior judgment of the CDLs. The right-hand panel displays the counterfactual and actual acquittal rates for Law graduates. The counterfactual acquittal rate is the same for Law graduates and for other defendants (22.4% vs. 23.7%). Accordingly, based on the evidence of participation in the resistance available to the *Jury* and on the initial decision of the CDLs, the acquittal rate for Law graduates should have been the same as that of other defendants. Yet, the actual acquittal rate of Law graduates is 30.9%. This is 8.5 percentage points higher than the counterfactual giving a strong indication that Law graduates enjoyed a particular advantage before the *Jury*.

[20] Appendix B.1 provides details.

Figure 1: The two courts' decisions - Law graduates vs. others defendants

(a) Acquittal rates before the CDLs and the *Jury*
(Confidence intervals: 95%)



(b) Percentage of CDL decisions overruled by the *Jury*
(Confidence intervals: 95%)

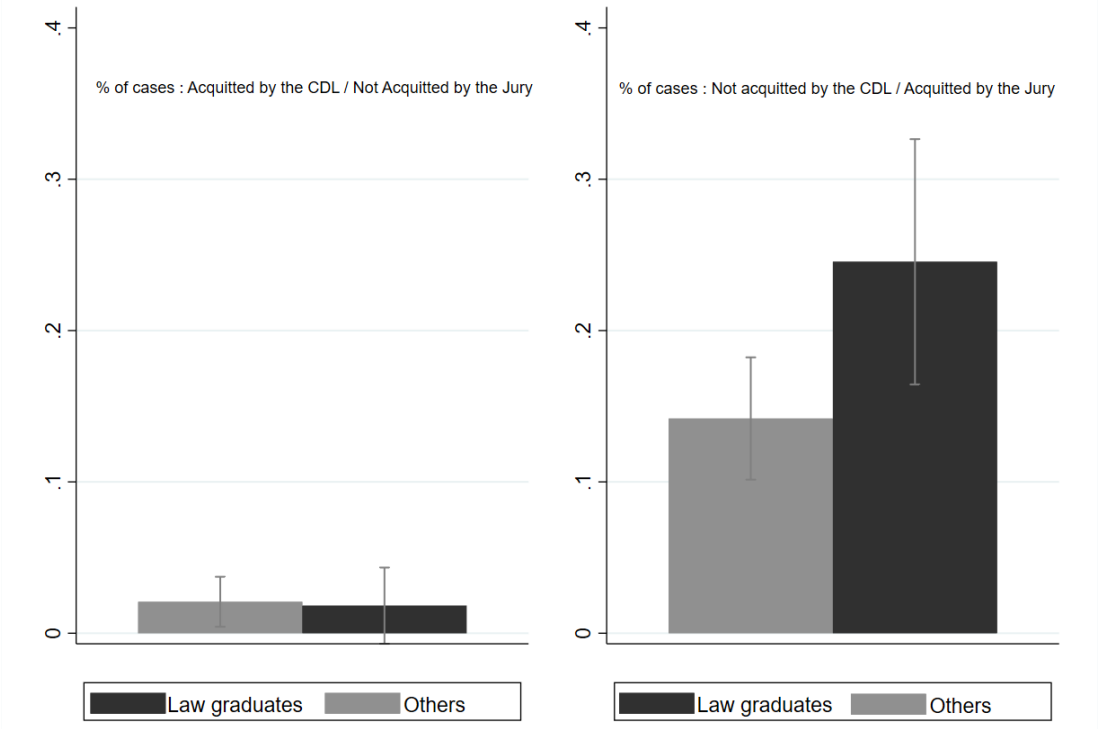
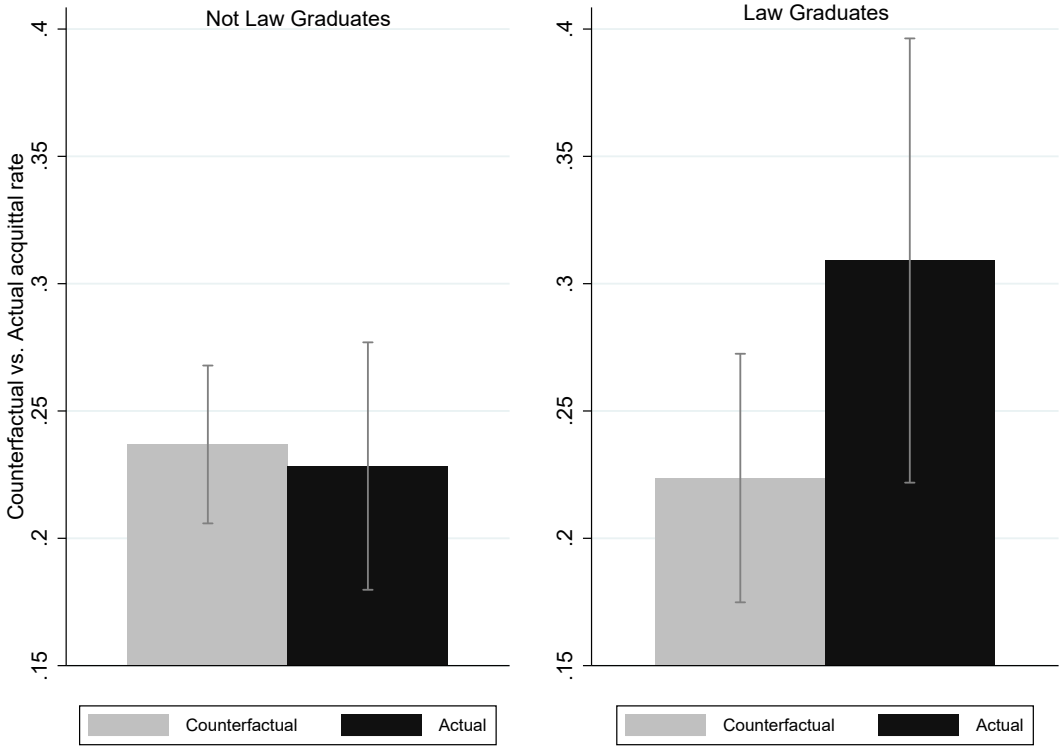


Figure 2: Counterfactual vs. actual acquittal rates - Law graduates vs. other defendants



4.2 Empirical strategy

The evidence presented above suggests that Law graduates were more likely to be acquitted by the *Jury* than other defendants. However, this advantage could, as in our theory (Section 3), be due to statistical-discrimination or the possible advantages provided by other characteristics correlated with being a Law graduate. The objective of this section is to causally identify the Law graduate advantage before the *Jury*. To do this end, we leverage the fact that each defendant was tried twice – first, by a CDL and, then, by the *Jury* – and that he was either a Law graduate or not to develop a difference-in-differences identification strategy. Let $Acquit_{i,c}$ be a dummy variable equal to one if defendant i is acquitted by court $c \in \{L, U\}$, where $c = L$ indexes the CDLs and $c = U$ indexes the *Jury*. We set the dummy variable LG_i to one if defendant i is a Law graduate and zero otherwise, and the dummy variable $Jury_c$ to one if the judgment was reached by the *Jury* and zero if the judgment was reached by a CDL. We estimate the following linear probability model:

$$Acquit_{i,c} = \alpha + \beta_1 Jury_c \times LG_i + \beta_2 Jury_c + \beta_3 LG_i + \beta_4 Jury_c \times X_i + \beta_5 X_i + \varepsilon_{i,c}, \tag{11}$$

where α, β_1, β_2 and β_3 are coefficients, and $\varepsilon_{i,c}$ is an error term. Equation (11) is estimated with ordinary least squares.[21] Standard errors are clustered at the defendant level.

Our approach is similar to the difference-in-differences estimation strategy that Anwar and Fang (2006) and Alesina and La Ferrara (2014) use to document a racial bias in judicial decisions in the USA. In our case, the treatment group consists of defendants who are Law graduates and the pre- and post-treatments are being judged by the CDLs and the *Jury*, respectively. To understand how we can use this strategy to causally identify the difference in the acquittal rate of Law graduates relative to other defendants in the judgments of the *Jury* relative to the CDLs, it is useful to consider each term in detail. Coefficient β_2 on the $Jury_c$ dummy variable captures the difference in acquittal rate across all defendants between the *Jury* and the CDLs. It captures all unobserved characteristics of the *Jury*, such as the weight it puts on type 1 relative to type 2 errors, that may result in it being, on average, more or less lenient than the CDLs and corresponds to b_c in Equation (5). Coefficient β_3 on the LG_i dummy variable captures the difference between the acquittal rate of Law graduates and that of other defendants regardless of which court rules. This corresponds to b_g in Equation (5) for the group of Law graduates. Coefficient β_1 on the interaction term $LG_i \times Jury_c$ is the coefficient of interest. It isolates the relative Law graduate advantage before the *Jury* by capturing the difference in the acquittal rate of the *Jury* from that of the CDLs specifically for Law graduates relative to other defendants after controlling for fixed defendant characteristics and unobserved court-specific characteristics. In other words, β_1 measures how much more inclined than the CDLs the *Jury* was to acquit Law graduates.[22]

Our theory suggests that the courts may, in different ways, base their bar for acquittal on observable group characteristics. Some of these may overlap or correlate with being a Law graduate. In Equation (11), we, therefore, control for a vector of individual (group) characteristics (X_i) and their interactions with the *Jury* dummy variable ($Jury_c \times X_i$) to ensure that the estimate of β_1 does not capture any other group

[21] Results are similar with a Probit or Logit estimator (Appendix B.4). We chose a linear probability model as the baseline model since this makes the interpretation of the interaction effect straightforward.

[22] In the literature on sentence bias, this effect is referred to as 'taste-based discrimination' (e.g., Alesina and La Ferrara, 2014). We prefer to use the term "Law graduate advantage" and avoid the term discrimination.

characteristics correlated with the Law graduate dummy variable and with higher acquittal rates before the *Jury*. We investigate a large number of potential controls and retain those that are statistically significantly correlated with the difference in acquittal rates between the *Jury* and the CDLs.[23] We also include in X_i information on participation in the resistance or of collaboration with the Vichy regime obtained from the dossiers of the defendants and coded by Wieviorka (2001).[24] Controlling for these factors and their interactions with the *Jury* dummy variable ensures that our estimate of the Law graduate advantage is not driven by the fact that the *Jury* had better access to evidence of participation in the resistance than the CDLs or that Law graduates due to their legal training were better at conveying information on their participation in the resistance to the *Jury*.

Unlike most studies of court decisions, we face no "selection-into-encounters" problem (Knox et al., 2020) because the exact same population of defendants automatically faced the two courts within a short period of time. Suspicion and selection into a second court hearing (here by the *Jury*) can, therefore, not correlate with defendant characteristics. Likewise no selection into the purge process could occur because all parliamentarians who had voted in favor of the 1940 enabling act were automatically selected.[25]

4.3 Difference-in-differences results

Table 2 reports the estimates of Equation (11). Column 2.1 shows a parsimonious specification with the Law graduate and *Jury* dummy variables only. The difference-in-differences estimate of β_1 is significant at the five-percent level. The point estimate implies that the difference in acquittal rates between Law graduates and other defendants is 10,6 percentage points higher before the *Jury* than before the CDLs. The estimate of β_1 remains statistically significant and its magnitude is stable – between between 9.6 and 11.0 percentage points – across the specifications reported in the other columns, where we add controls for the characteristics of the defendants and the interaction between these characteristics and the *Jury* dummy variable. In particular, the estimate is hardly affected when we control for the presence of evidence of either collaboration with the Vichy regime or of participation in the resistance movement (Column 2.5). Additionally, the specification in Column 2.7 controls for defendant fixed effects, which neither affect the magnitude nor the significance of the estimate of β_1.

Table 2, therefore, shows that Law graduates were about 10 percentage points more likely to be acquitted by the *Jury* relative to the CDLs than other defendants. We interpret this effect as the Law graduate advantage before the *Jury,* but, as noted above, it could also reflect a disadvantage before the CDLs. To substantiate our interpretation, we study the decisions of the two courts separately and estimate the difference in acquittal

[23]The results are reported in Table B.2. The following defendants (group) characteristics were treated differently by the two courts and included as controls: age, being Jewish, being a journalist, being a President/Vice President or Secretary of the National Assembly, and a dummy equal to one if the defendant's constituency was in the part of France that was initially occupied by Germany. Conversely, the difference in acquittal rates before the CDLs and before the *Jury* is not explained by length of academic studies, longer political career, or political orientation (see the bottom of Table B.2).

[24]This includes dummy variables equal to one if there was proof of participation in the civilian or military resistance, if the defendant had been arrested by the Vichy regime, or if he had been a mayor under that regime.

[25]Admittedly, there is some attrition in our data, but it is random and should not affect our results. Out of the 569 parliamentarians who voted in favor of the 1940 enabling act, 93 died during the war, 51 had been acquitted by prefects prior to the legal process because they were well-known figures of the resistance, eight were facing criminal courts for collaboration with the Vichy regime, eight were from overseas and we, therefore, do not have CDL judgments in their dossiers. Finally, in the dossiers of 16 defendants the decision of either the CDL or the *Jury* is missing. Those cases were, typically, cases on which the *Jury* and CDLs would have agreed given the level of evidence of participation in the resistance or were lost randomly. Among those missing cases, the proportion of Law graduates is not statistically different from what it is for the defendants who are not missing (32% vs. 28%).

Table 2: The advantage of Law graduates before the *Jury*: Difference-in-differences estimates

	(2.1)	(2.2)	(2.3)	(2.4)	(2.5)	(2.6)	(2.7)
Dep variable	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$
Estimator	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Jury	0.121***	0.419***	0.106***	0.218***	-0.0224	0.143	0.143
	(5.336)	(3.283)	(4.503)	(5.000)	(-0.887)	(1.278)	(1.287)
LG	-0.0254	-0.0264	-0.0276	-0.0273	-0.0249	-0.0284	
	(-0.797)	(-0.839)	(-0.856)	(-0.831)	(-0.813)	(-0.915)	
Jury × LG	0.106**	0.0978**	0.110**	0.0958*	0.108**	0.0996**	0.0996**
	(2.143)	(1.985)	(2.217)	(1.890)	(2.385)	(2.149)	(2.165)
Constant	0.107***	0.0794	0.116***	0.101***	0.0498**	-0.0797	0.100***
	(5.874)	(0.907)	(5.773)	(3.405)	(2.197)	(-0.829)	(10.87)
Controls:							
Age and Religion		Yes				Yes	Yes
Journalist			Yes			Yes	Yes
Political mandates				Yes		Yes	Yes
Resistance and collaboration WWII					Yes	Yes	Yes
Individual FE							Yes
Observations	798	798	798	798	798	798	798
Adjusted R-squared	0.040	0.052	0.041	0.049	0.271	0.281	0.289

Note: Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Jury is a dummy variable equal to one if the judgment was before the Jury and zero if it was before a CDL. LG is a dummy variable equal to one if the defendant was a Law graduate and zero otherwise. This table presents estimates of Equation (11 . Controls are: Age and religion (Age, Jewishness); Journalist; Political mandates (Mayor, Special Role in the Assembly, parliamentarian of an occupied territory); Resistance and collaboration WWII (Civilian Resistance, Military resistance, Arrested by the Vichy regime, Mayor under Etat Francais). Each control is also interacted with the Jury dummy variable.

rates for Law graduates relative to other defendants for each (see Appendix B.3). The results show that the Law graduate advantage comes from differences in sentencing patterns before the *Jury* and not before the CDLs: Law graduates had a 7.1 percentage points *higher* probability than other defendants to be acquitted by the *Jury* and a 2.8 percentage points *lower* probability to be cleared by a CDL. Accordingly, around 71 percent of the relative Law graduate advantage appears before the *Jury* and the effect is mostly driven by Law graduates whom the CDLs wanted to ban but the *Jury* decided to acquit (see Appendix B.4). This is in line with the descriptive evidence from Figure 1(a). The baseline difference-in-differences estimates reported in Table 2 are robust to alternative estimations methods (Appendix B.4), to controlling for the composition of resistance groups in a département, which we interpret as a proxy of the political composition of the CDLs (Appendix B.5), and did not emerge because the CDLs disregarded evidence put forward by Law graduates (Appendix B.6).

In conclusion, Law graduates benefited from a 10 percentage points acquittal bonus before the *Jury* even after taking evidence of participation in the resistance and decisions of lower courts into account. Accordingly, at least 11 of them might have survived the purge because of the Law graduate advantage. This is a sizable effect equal to the size of the smallest parliamentary groups in the French Parliament in 1946.

4.4 The Law graduate advantage and elite persistence

For the Law graduate advantage to have contributed to elite persistence, it must have impacted the future careers of the politicians who benefited from it. The advantage would have been inconsequential if acquitted Law graduates never ran for election or if voters did not elected them. In this section, we show that the decisions of the *Jury* were consequential for the post-war careers of defendants and that Law graduates were

more likely than other defendants to pursue a career in politics after being cleared by the *Jury*.

Table 3 reports a series of regressions in which measures of each defendant's post-war political career (such as the number of times he ran in municipal and legislative elections, whether he was mayor, held a seat in parliament or had ministerial responsibilities) are explained by the court decisions or by being a Law graduate. In each regression, we control for the CDLs' decision to acquit the defendant to capture the deviation from the CDLs' decisions, in the spirit of our difference-in-differences estimation strategy. Panel A assesses the impact of acquittal by the *Jury* on the defendant's post-war political career. If those acquitted became politically inactive, we should observe a null result. Instead, we see that defendants acquitted by the *Jury* ran frequently in municipal and legislative elections (Column 3.A.1 and 3.A.2). As a result, they were 12 percentage points more likely to be a mayor (Column 3.3), 23 percentage points more likely to hold a seat in parliament (Column 3.A.4), and seven percentage points more likely to become a government minister than defendants who were banned (Column 3.A.5). Panel B assesses the political careers of Law graduates after the war relative to other defendants. The dependent variables in the regressions are the same as in Panel A, but the explanatory variable of interest is the Law graduate dummy. We see that Law graduates were more active in national politics than other defendants. They ran more frequently in national legislative elections (Column 3.B.2) and, as a result, they were around eight percentage points more likely to become a deputy (Column 3.B.4) and three percentage points more likely to hold ministerial responsibilities (Column 3.B.5) than other defendants.[26] In contrast, Law graduates were not more active or successful in local politics than other defendants (Column 3.B.1 and 3.B.3). Finally, in Appendix B.7, we estimate Equation (11) separately for defendants who ran for election in the first post-war local elections in 1945 and those who did not. We find that the Law graduate advantage is present only for defendants who ran and who in that way revealed their ambition to stay in politics after the war.

In conclusion, the difference-in-differences analysis from Section 4.3 shows that Law graduates had an advantage that helped them avoid the post-war purge in the transition back to democracy: they were about ten percentage points more likely to be acquitted than other defendants accused of collaborating with the previous autocratic regime. This section shows that Law graduates were more likely than other defendants to be active and successful in post-war national politics. Together this shows that group-specific advantages in a democratic purge process can lead to elite persistence – increasing the probability of some elite persisting in comparison to others. The key question remaining is: what was the mechanism behind the Law graduate advantage? Section 5 answers this question by studying the content of the defendants' dossiers.

[26] The results are very similar if we measure political success with the number of years a cleared defendant spent in the different positions (see Appendix B.8)

Table 3: Elite persistence - Jury decisions and Law graduates

Panel A - Independent variable of interest: Acquitted$_{Jury}$

	(3.A.1) Nb mayor election	(3.A.2) Nb deputy elections	(3.A.3) Mayor=1	(3.A.4) Parliament=1	(3.A.5) Minister=1
Acquitted$_{Jury}$	0.189**	0.253***	0.116**	0.225***	0.0701**
	(2.550)	(3.458)	(2.210)	(3.787)	(2.293)
Acquitted$_{CDL}$	0.0431	0.0735	0.0115	0.0553	0.00894
	(0.469)	(0.739)	(0.180)	(0.780)	(0.228)
Constant	0.595***	0.630***	0.379***	0.349***	0.0720*
	(4.590)	(4.594)	(3.842)	(3.609)	(1.681)
Observations	397	399	399	399	399
Adj R^2	0.172	0.156	0.166	0.157	0.067
Control variables					
Individual	YES	YES	YES	YES	YES

Panel B - Independent variable of interest: Law graduate

	(3.B.1) Nb mayor elections	(3.B.2) Nb deputy elections	(3.B.3) Mayor=1	(3.B.4) Parliament=1	(3.B.5) Minister=1
Law graduate	-0.0352	0.105**	-0.0107	0.0761**	0.0318*
	(-0.836)	(2.210)	(-0.348)	(2.103)	(1.691)
Acquitted$_{CDL}$	0.121	0.188*	0.0597	0.156**	0.0408
	(1.303)	(1.953)	(0.926)	(2.162)	(1.072)
Constant	0.630***	0.626***	0.397***	0.351***	0.0699
	(4.789)	(4.554)	(4.023)	(3.654)	(1.513)
Observations	397	399	399	399	399
Adj R^2	0.153	0.125	0.151	0.110	0.044
Control variables					
Individual	YES	YES	YES	YES	YES

Note: Level of observation: Defendant. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1. In Panel A, Acquitted$_{Jury}$ is a dummy variable equal to 1 if the defendant was acquitted by the *Jury*. In Panel B, Law gradaduate is a dummy variable equal to one if the defendant was a Law graduate. Acquitted$_{CDL}$ is a dummy variable equal to one if the defendent was acquitted by a CDL. Dependent variables: Nb mayor elections is the number of times a defendant was a candidate in a municipal election after the war. Nb deputy elections is the number of times a defendant was a candidate in a legislative election after the war. Mayor=1 is a dummy variable equal to one if a defendant was elected mayor after the war and zero otherwise. Parliament=1 is a dummy variable equal to one if a defendant was a deputy or a senator after the war and zero otherwise. Minister=1 is a dummy variable equal to one if the defendant was a minister after the war and zero otherwise. Individual controls include: Age and religion (Age, Jewishness); Journalist; Political mandates (Mayor, Special Role in the Assembly, parliamentarian of an occupied territory); Resistance and collaboration WWII (Civilian Resistance, Military resistance, Arrested by the Vichy regime, Mayor under Etat Francais).

5 What explains the Law graduate advantage?

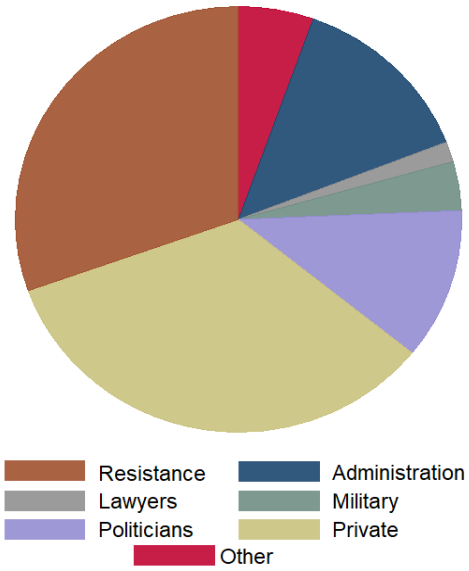
In this section, we leverage the content of the dossiers of the individual defendants kept in the archives of the *Jury* to investigate the role of connections in generating the Law graduate advantage. We begin by presenting the data set built from the individual dossiers and then test if, as suggested by our theory, direct and indirect connections between the defendants or their supporters and the *Jury* can explain the Law graduate advantage. We conclude by ruling out that the advantage is caused by legal skills.

5.1 The dossier data set

We created a full inventory of the documents contained in the individual dossiers of the defendants facing the *Jury*.[27] The *Jury* kept a detailed record of each case, including all internal and external correspondence, and the dossiers contain all pieces of information the *Jury* used to reach its ruling. Overall, the inventory pertains to 17,589 documents. The inventory makes it possible to extract and quantify three types of information. First, we can quantify the volume and length of different types of documents in each of the dossiers and classify them according to their content. Second, we can record communication between defendants and the *Jury*: direct connections. Third, we can record letters from individuals external to the *Jury* who tried to intervene in favor or against a defendant by writing to the *Jury* or to third parties connected to the *Jury*: indirect connections. We refer to these as "letters of support".[28] From the dossier data set, we code variables measuring for each defendant these three aspects. Appendix Table C.2 defines and presents summary statistics of the variables related to the content of the dossiers used in the analysis whereas Appendix Table D.2 reports summary statistics related to the overall structure of the dossiers.

Figure 3 presents information on the origin of letters of support. These letters represent 19.2% of the documents in a typical dossier.[29] A third of them were private correspondence sent by the defendant's friends, family, or by individuals in his constituency ("private") and 30.5% originated from resistance organizations.

Figure 3: The origins of letters of support in the dossiers of the defendants



Note: The categories are defined as follows. Administration refers to documents produced by a ministry or a local administration. Vichy Regime refers to documents from the archives of the Vichy regime or from any Vichy-related institutions. Defendant refers to documents produced by the defendant himself. Jury refers to documents produced by the *Jury*. Military refers to documents produced by the French army. Private refers to documentd produced by an individual in his/her own name without stating an obvious relation to an organization typically coming from family members or friends of the defendant or from individuals in his constituency. Politicians refers to documents produced by parliamentarians and local politicians. Resistance refers to documents produced by members of resistance networks. Lawyers refers to documents sent by a lawyer using his/her title in the document sent.

Table 4 compares the dossiers of Law graduates with those of other defendants. The top panel reports mean comparisons for variables related to the structure of the dossiers; the middle panel reports comparisons for

[27] The dossiers are to be found in the French National Archives (References AL//5295 to AL//5334).

[28] We note that in a few cases, opponents submitted letters in support of purging the defendant, but the vast majority of the letters were, in fact, letters advocating that the defendant be acquitted.

[29] Appendix C.1 presents information on the origin of all documents in the dossiers.

variables related to letters of support; and the bottom panel reports comparisons for variables related to the information content of the dossiers. We see that the dossiers of Law graduates and other defendants are strikingly similar and in all, but one aspect, the means are statistically indistinguishable.[30] In short, the dossiers of Law graduates are not "thicker" than other dossiers and do not contain more letters of support or more letters with information about participation in the resistance than those of other defendants.

Table 4: Law graduates and the content of dossiers

	(4.1)	(4.2)	(4.3)
		Mean	
	LG	Others	Diff=0 (p-value)
Structure of the dossiers			
Nb Documents	40.96	40.59	0.90
Nb Pages	59.30	53.97	0.29
Nb Documents from the *Jury*	13.86	13.63	0.74
Nb Archival Documents	3.22	2.63	0.39
Nb Information requests	1.16	1.04	0.27
Letters of support			
Nb Letters of support	7.16	8.06	0.52
Nb Letters of support - in Favor	6.78	7.65	0.52
Nb Letters of support - Against	0.22	0.22	0.98
Nb Letters of support - Neutral	0.16	0.19	0.74
Nb Letters of support - Resistance	2.09	2.54	0.35
Nb Letters of support - Military	0.24	0.28	0.72
Nb Letters of support - Administration	1.22	0.94	0.39
Nb Letters of support - Others	2.08	2.88	0.26
Information content			
Nb Documents - Military resistance	0.82	0.75	0.85
Nb Documents - Civilian resistance	8.36	8.63	0.83
Nb Documents - Resistant Press	0.29	0.70	0.04**
Nb Documents - Legal Arguments	5.59	5.37	0.62
Nb Documents - Political opinion	10.39	9.87	0.75
Nb Documents - Reelection	0.95	1.11	0.53
Nb Documents - Other topic	6.45	6.26	0.84

Note: 'Nb' means number of. The top panel, labelled Structure of the dossiers, presents statistics on the overall number of documents in the dossiers and the number of documents of various types. The middle panel, labelled "Letters of support", presents statistics on letters of support broken down in subcategories. The first subcategory relates to the opinion expressed in these letters (in favor, neutral, against). The second subcategory relates to the affiliation of the sender (Resistance, Military, Administration). The bottom panel, labelled information content, presents information on the topic covered by the documents. For example, the first line of this panel headed Nb Doc - Military resistance should be read as: the average number of documents providing information on actions related to participation in military resistance. Columns 4.1 and 4.2 display the mean value for the group of Law graduates (LG) and other defendants (Others), respectively. Column 4.3 reports the p-value of a two-sided t-test of equal means with *** p<0.01, ** p<0.05, * p<0.1.

5.2 The Law graduate advantage and connections: empirical strategy

To test if connections can account for the Law graduate advantage, we augment Equation (11) and estimate the following:

[30]The exception is documents referring to the "Resistant Press", which are less common in the dossiers of Law graduates. This is probably due to the fact that Law graduates are rarely journalists. In any case, that difference cannot explain the Law graduate advantage because we control for that difference in the difference-in-differences estimations reported in Table 2.

$$Acquit_{i,c} = \alpha + \beta_1 Jury_c \times LG_i + \beta_2 Jury_c + \beta_3 LG_i + \Gamma_1(Court_c + Defendant_i) \times C_i + \beta_4 Jury_c \times NbDoc_i + \beta_5 NbDoc_i + \varepsilon_{i,c}, \quad (12)$$

where $Acquit_{i,c}$ is the dummy variable equal to one if defendant i is acquitted by court c and the rest of the first line corresponds to the baseline specification in Table 2, Column 1. The terms related to connections are shown in the second line along with the error term $\varepsilon_{i,c}$. The matrix $(Court_c + Defendant_i)$ includes four dummy variables defined for pairs of defendants (Law graduate or another defendant) and court (*Jury* or CDL). These dummy variables are interacted with C_i which is a measure of either direct or indirect connections for a given defendant *i*. Γ_1 is the vector of coefficients measuring the effect of connections for different subcategories of decisions defined by the court/defendant-type pairs. $NbDoc_i$ is the number of documents in the dossier of defendant *i*. It is included directly and interacted with the Law graduate dummy variable to control for the possibility that the volume of document in the dossiers can influence the difference in acquittal rates between the CDLs and the *Jury*.

Equation (12) allows us to test if connections can account for the Law graduate advantage in two ways. First, we can establish if a measure of connections explains the Law graduate advantage by testing if the coefficient associated with the interaction $Jury_c + LG_i$ turns insignificant when that measure is controlled for. Second, we can test if connections only worked before the *Jury* and for Law graduates, as opposed to before the CDLs and/or for other defendants. If so, only the interaction $(Jury + LG) \times C_i$ will be significant and a Wald-test will indicate that the coefficients on the interactions between the four court/defendant-types dummy variables and the measure of connections are statistically different.

5.3 Can direct connections explain the Law graduate advantage?

We begin by assessing if direct connections can explain the Law graduate advantage. We measure direct connections by the number of documents presented by the defendant himself and the type of communication between the defendant and the *Jury* contained in the dossiers. We distinguish between communication related to legal arguments and communication related to non-legal matters, e.g., letters asking for an update on how the case is proceeding, etc. First, we conjecture that defendants with lower cost of communication will communicate more frequently with the *Jury* about matters related to their case than others. Second, we conjecture that defendants who write in an informal style are better connected to the *Jury*. We measure informality in the letters sent by defendants through the greeting used and count the number of letters addressing the recipient as "Dear X" in a dossier. The use of "dear" in French signals a degree of informality and therefore a tighter link between the sender and the recipient. We refer to such documents as "Informal letters from defendant".

Table 5: Direct connections between the defendants and the *Jury*: Law graduates vs. other defendants

	(5.1)	(5.2)	(5.3)
	Mean		
	LG	Others	Diff=0 (p-value)
1. Nb Documents from defendant	5.00	4.50	0.46
1.1 Nb Documents - Communication from defendant without legal content	1.80	1.32	0.05*
1.2 Nb Documents - Communication from defendant with legal content	3.2	3.2	0.98
1.3 Nb pages - Communication from defendant without legal content	2.62	1.59	0.02**
1.4 Nb pages - Communication from defendant with legal content	11.6	8.95	0.28
2. Nb Informal letters from defendant	0.17	0.16	0.89
2.1 Nb Documents - Informal letters without legal content	0.08	0.05	0.32
2.2 Nb Documents - Informal letters with legal content	0.09	0.11	0.68

Note: 'Nb' means number of. In row 1, "Nb Documents from defendant" refers to the number of documents sent by the defendant to the *Jury*. In row 1.1, "Nb Documents - Communication from defendant without legal content" refers to the number of documents sent by the defendant where the content is not directly related to the legal aspect of his case. In row 1.2, "Nb Documents - Communication from defendant with legal content" refers to the number of documents sent by the defendant in which the defendant presents legal information related to his case. In row 1.3, "Nb pages - Communication from defendant without legal content" counts the pages in "Nb Documents - Communication from defendant without legal content". In row 1.4, "Nb pages - Communication from defendant with legal content" counts the number of pages in "Nb Documents - Communication from defendant with legal content". In row 2, 'Nb Informal letters from defendant' refers to the number of informatl letters in the dossiers; in row 2.1 and 2.2, this is split between information letters with and without legal content. Columns 5.1 and 5.2 display the mean value for the group of Law graduates ('LG') and other defendants ('Others'), respective. Column 5.3 presents the p-value associated with a two-sided t-test for equal means. The level of significant is indicated with *** p<0.01, ** p<0.05, * p<0.1.

The top panel of Table 5 reports mean comparisons between Law graduates and other defendants for direct connections as measured by the volume of communication between the defendant and the *Jury*. Row 1 shows that Law graduates, on average, sent 5.0 letters directly to the Jury, while other defendants, on average, sent only 4.5, but the difference is not statistically significant. Thus, it does not appear that Law graduates communicated more with the *Jury* by sending more letters in total than other defendants. However, the total number of letters masks differences in the type of arguments presented in them. In rows 1.1 to 1.4, we break down the letters into those with legal arguments and those without. On the one hand, Law graduates submitted more and longer letters without legal arguments than other defendants (rows 1.1 and 1.3). Specifically, they sent on average an extra page of communication not directly related to the legal aspects of their case to the *Jury* (row 1.3). On the other hand, the difference in the length and number of letters with legal arguments is not statistically significant (rows 1.2 and 1.4). The bottom panel reports mean comparisons for the other measures of direct connections related to the degree of informality in the letters. We find no statistical difference between Law graduates and other defendants (rows 2, 2.1 and 2.2). To summarize, the evidence suggests that Law graduates had direct connections to the *Jury* but did not use these connections more than other defendants to defend themselves using legal arguments. Instead, they used them to communicate with the *Jury* about non-legal aspects of their case.

To test if these measures of direct connections can account for the Law graduate advantage, we estimate Equation (12) with interactions between the four court-defendant-type pair dummy variables and the measures of direct connections. Table 6 reports the results.

Table 6: Direct connections as the origin of the Law Graduate advantage

	(6.1)	(6.2)	(6.3)
Dep Variable	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$
Sample	All	All	All
Measure of Connections C_i	Nb Docs from Def	Nb Doc - Com Def without legal arg	Nb Pages - Com Def without legal arg
Jury × LG	0.158**	0.176***	0.176***
	(1.985)	(2.832)	(2.769)
(Jury + LG) × C	0.0491	0.0141	0.0108
	(0.779)	(0.208)	(0.189)
(Jury + Others)× C	0.0375	0.0567	0.0463
	(0.951)	(1.300)	(1.196)
(CDL + LG)× C	0.0401	0.0583	0.0525
	(0.920)	(1.341)	(1.311)
(CDL + Others)× C	-0.0151	-0.0164	-0.0181
	(-0.598)	(-0.613)	(-0.764)
Nb of Docs as control	Yes	Yes	Yes
Wald-test equality interactions	0.43	0.15	0.15
Observations	798	798	798
Adjusted R-squared	0.055	0.056	0.056

Note: Estimates of Equation (12) with OLS controlling for the number of documents in each invidual dossier and its interaction with the Jury dummy variable. The specifications include interactions between the four dummy variables identifying "Court-defendant-type" pairs and a particular measure of direct connections C_i as indicated in the collumn headings. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The regression results show that direct connections to the *Jury* do not explain the Law graduate advantage. Firstly, the coefficient of the interaction $Jury_c \times LG_i$ remains significant at usual levels and its magnitude increases from 0.10 to 0.17 after we control for the effect of direct connections to the courts. Secondly, the coefficients on the measures of direct connections interacted with the four court-defendant-type dummy variables are all statistically insignificant.

In conclusion, although Law graduates had more direct connections to the *Jury* than other defendants and communicated more with it and on matters unrelated to the legal aspects of their defense, these direct connections do not explain why Law graduates enjoyed an advantage before the *Jury*. Viewed through the lens of our model this suggests that that Law graduates were too connected to the *Jury* and that their non-legal arguments were, therefore, discounted. The next section investigates if the Law graduate advantage, instead, can be explained by indirect connections.

5.4 Can indirect connections explain the Law graduate advantage?

In this section, we build three measures of indirect connections between the supporters of a defendant and the *Jury* and investigate if they can explain the Law graduate advantage. First, we measure indirect connections by the number of letters of support in a defendant's dossier originating from supporters in Paris and by the number of different Paris-based supporters. The rationale is that the *Jury* and its members were located in Paris and that the three judges had studied Law at Parisian Universities. We conjecture that Paris-based authors of letters of support, therefore, had a stronger connection to the *Jury* than other authors, as they would interact in the same social and professional circles. Second, we built an index of the portfolio of letters of support of each defendant where the letters in the portfolio are weighted by how connected to

the *Jury* the supporters sending them were. To do this, we divide the letters of support in a defendant's dossier into groups where letters originated from the same "organization" (such as, for example, a ministry, a prefecture, or a bureau) are grouped together and calculate the fraction of letters from each organization.[31] We index those organizations with $s = 1, ..., n$ where n is the total number of organizations in the data set. We conjecture that a supporter has a stronger indirect connection to the *Jury* if he is associated with an organization that sends more documents and letters of support to the *Jury* in general, i.e., in relation to all cases. For example, the Ministry of Home Affairs corresponded more with the *Jury* than the Prefecture of the Morbihan *département*. We hypothesize that a letter from someone in the Ministry of Home Affairs would carry more weight than a letter from someone from the Prefecture of the Morbihan *département*. We, therefore, weight the share of letters in a defendant's portfolio from organization s with a measure of the total volume of correspondence originating from that organization in the entire data set. Formally, for defendant i, the index is defined as

$$\sum_{s=1}^{n} \left(\frac{NbLetters_{s,i}}{\sum_{s=1}^{n} NbLetters_{s,i}} \times weight_s \right), \tag{13}$$

where $\frac{NBLetter_{s,i}}{\sum_{s=1}^{n} NbLetters_{s,i}}$ is the share of letters of support in defendant $i's$ portfolio that originated from supporters associated with organization s and $weight_s$ is the weight attached to organization s.[32] We calculate two different versions of this index. The first version weights the share with the total number of documents ($NbDocs_s$) from organization s across all defendants (i.e., $weight_s = NbDocs_s$) and is called "Indirect connections via supporters". The second version excludes all letters of support ($NbLetters_s$) from the weight attached to each organization (i.e., $weight_s = NbDocs_s - NbLetters_s$). The aim is to capture the connection between supporters of a defendant i and the *Jury* net of advocacy in support of defendants in general. For example, if the Ministry of the Interior sent X letters of supports (with arguments for why a defendant should or should not be banned) and Y administrative reports with factual information, then the weight is only based on Y. We call this measure "Indirect connections via supporters (excl. Letters)".

Third, we record the number of documents from supporters of each defendant addressing the recipient as "Dear X" and use these "informal documents from supporters" as an index of his indirect connections.

Table 7 reports mean comparisons of these proxies for indirect connections for Law graduates (LG) and other defendants ($Others$). The results in rows 1 and 2 show that Law graduates received more letters of support from supporters in Paris and that they had more Paris-based supporters than other defendants.[33] Moreover, the two indexes of the portfolio of letters of support also indicate that the supporters of Law graduates were better connected to the Jury than those of other defendants (rows 3 and 4). Finally, the dossiers of Law graduates contained more "informal documents from supporters" than those of other defendants (row 5). The difference is significant for letters addressed to third parties to the case (row 5.3), but not for documents directly sent to the *Jury* (rows 5.1 and 5.2). Overall, the evidence presented in Table 7 shows that

[31] We classify "organizations" along two dimensions: its name (e.g., Ministry of Home Affairs) and its location (e.g., Paris). For example, a bureau of the Ministry of Home Affairs located in Lyon is considered as a different entity than the Ministry itself located in Paris. Sometimes, supporters do not belong to any specific organization: so for each *département* there is a fictitious "organization" of people not affiliated to any organization ("individuals"). Among organizations, individuals from Paris, the Ministry of Home Affairs (Paris), the National Assembly (Paris), the Prefecture of the Seine *département* (Paris), the Prefecture of the North *département*, the Prefecture of the Morbihan *département*, and individuals from the Nièvre *département* are the ones that are more "connected" to the *Jury*.

[32] In cases where a defendant did not get any letters of support, the index takes the value of 0.

[33] The two groups have the same total number of letters of support (see Table 4). Accordingly, it is the composition of who sent them that differs.

in the dossiers of Law graduates there were more letters coming from well-connected supporters and that Law graduates were able to leverage connections to third parties to seek influence the *Jury*. In short, Law graduates were connected to supporters with stronger indirect connections to the *Jury* than other defendants.

Table 7: Indirect connections between the defendants and the *Jury*: Law graduates vs. other defendents

	(7.1)	(7.2)	(7.3)
	Mean		
	LG	Others	Diff=0 (p-value)
1. Nb Letters of support from Paris	2.72	1.82	0.09*
2. Nb Supporters from Paris	1.23	0.70	0.01***
3. Indirect connections via supporters	47.41	27.82	0.03**
4. Indirect connections via supporters (excl. letters)	26.59	15.13	0.03**
5. Nb Informal documents	1.99	1.26	0.03**
5.1 Nb Informal documents to Cassin	0.15	0.20	0.47
5.2 Nb Informal documents to Jury	0.22	0.22	0.96
5.3 Nb Informal documents not to Jury	1.77	1.04	0.02**

Note: In row 1, "Nb Letters of support from Paris" refers to the number of letters of support with a sender located in Paris. In row 2, "Nb Supporters from Paris" refers to how many different supporters from Paris were sending letters of support. In row 3, "Indirect connections via supporters" refers to the portfolio index based on the total number of documents from each organization. In row 4, "Indirect connections via supporters (excl. letters)" refers to the portfolio index based on the total number of documents from each organization net of letters of support. In row 5, "Nb Informal documents from supporters" refers to the number of documents including "Dear" in their headings. Rows 5.1 to 5.3 decompose the total number of informal documents from row 5 depending on the sender and receiver of such documents. Columns 7.1 and 7.2 display the mean value for the group of Law graduates (LG) and other defendants (Others), respectively. Column 7.3 reports the p-value of a two-sided t-test of equal means and its level of significance with *** p<0.01, ** p<0.05, * p<0.1.

To investigate if these observed differences in indirect connections can explain the Law graduate advantage, we estimate Equation (12) and test if controlling for the various measures of indirect connections interacted with the four court/defendant-type dummy variables eliminates the Law graduate advantage. Table 8 reports the results. We observe that the interaction $Jury_c \times LG_i$ is insignificant in three specifications: when indirect connections are measured by letters of support from Paris (column 8.1) or by the two indexes of indirect connections (columns 8.3 and 8.4). We can, accordingly, infer that these three measures of indirect connections at least partially explain the Law graduate advantage before the *Jury*. In contrast, in columns 8.5 and 8.6, the size of the coefficients on $Jury_c \times LG_i$ remain close to those of the baseline estimates and the coefficients are significant at the five percent level. This suggests that indirect connections as inferred by informality in the communication between supporters of the defendants and the *Jury* cannot explain the Law graduate advantage. We also observe that the $Jury_c \times LG_i$ interaction remains marginally significant in the specification in column 8.2 where we proxy indirect connections with the number of different Paris-based supporters.

For the three cases in which controlling for indirect connections eliminates the Law graduate advantage, we can gain further insights by studying the interactions between these measures of indirect connections and the four court/defendant-type dummy variables. This is informative about the precise channel through which these connections operated. First, the results in column 8.1 and 8.4 show that letters of support from Paris-based supporters and "indirect connections via supporters (excl. letters)" mattered only for the decisions of the *Jury* in relation to Law graduates and not in the decisions of the *Jury* related to other defendants or in decisions by the CDLs. The result of the Wald-test reported at the bottom of the columns indicates that these differences are statistically significant in the case of the letters of support from Paris-based supporters. These results point to a specific effect of letters of support from Paris-based supporters for Law graduates in

front of the *Jury*. They suggest that the Law graduate advantage came from the ability of Law graduates to leverage Paris-based connections. Second, the interactions between "indirect connections via supporters" and the $(Jury + LG)$ and the $(Jury + Others)$ court/defendant-type pair dummies (column 8.3), respectively, are both positive and significant. Accordingly, the *Jury* was more likely to acquit all types of defendants when it received letters of support from organizations with which it frequently interacted. The Wald-tests reported at the bottom of the column show that we cannot reject that the coefficients on the four interaction terms are, in fact, the same.

Table 8: Indirect connections as the origin of Law graduate advantage

Dependent variable	(8.1) $Acquit_{i,c}$	(8.2) $Acquit_{i,c}$	(8.3) $Acquit_{i,c}$	(8.4) $Acquit_{i,c}$	(8.5) $Acquit_{i,c}$	(8.6) $Acquit_{i,c}$
Measure of Connections C_i =	Nb Letters from Paris	Nb Supporters from Paris	μ Indirect Connections	μ Indirect Connections (excl. letters)	Nb Doc Informal	Nb Doc Informal Not to Jury
Jury × LG	0.0766	0.0949*	0.0768	0.0707	0.123**	0.127**
	(1.518)	(1.915)	(1.376)	(1.318)	(2.115)	(2.216)
(Jury + LG) × C	0.119**	0.154**	0.0459**	0.0453*	0.0393	0.0372
	(2.374)	(2.291)	(2.003)	(1.800)	(0.660)	(0.618)
(Jury + Others) × C	0.0481	0.0813	0.0364**	0.0221	0.0803*	0.0829*
	(1.188)	(1.434)	(1.966)	(1.108)	(1.764)	(1.795)
(CDL + LG) × C	0.0377	0.0746	0.0146	0.0129	0.0222	0.0166
	(1.139)	(1.582)	(1.095)	(0.878)	(0.603)	(0.463)
(CDL + Others) × C	-0.0177	-0.0139	0.0111	0.00298	0.0112	-0.0118
	(-0.869)	(-0.496)	(0.825)	(0.213)	(0.368)	(-0.402)
Constant	0.137	0.171	0.206	0.164	0.169	0.130
	(1.014)	(1.378)	(1.550)	(1.308)	(1.314)	(1.033)
Control Nb Docs	Yes	Yes	Yes	Yes	Yes	Yes
Control Nb Docs X Jury	Yes	Yes	Yes	Yes	Yes	Yes
F-test equality of Interactions	0.03**	0.05**	0.28	0.39	0.53	0.23
Observations	798	798	798	798	798	798
Adjusted R-squared	0.065	0.066	0.064	0.060	0.058	0.058

Note: Estimates of Equation (12) with OLS controlling for the size of each invidual dossiers and its interaction with the Jury dummy variable. Each column includes the interaction between the four court/defendant-type dummy variablesand C_i. The relevant measure of indirect connections C_i is defined in the heading to each column. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

In conclusion, letters of support from Paris-based supporters specifically benefited Law graduates. Indirect connections to the *Jury,* as measured by our two portfolio indexes, helped all types of defendants getting acquitted, but Law graduates were more frequently supported by third parties with such connections to the *Jury* (Table 7). These results suggest that the Law graduate advantage is, at least in part, caused by indirect connections.

5.5 Can the Law graduate advantage be explained by legal skills?

In the previous subsections, we have established that Law graduates had better direct and indirect connections to the *Jury* than other defendants (Tables 5 and 7). While direct connections cannot explain the Law graduate advantage before the *Jury* (Section 5.3), we find consistent evidence that indirect connections can (Section 5.4).

There is, however, an alternative explanation that we need to rule out: the Law graduate advantage could simply be due to their legal skills. Although the dossiers of Law graduates, as documented in Table 4, did not look different from those of other defendants (except with regard to connections) and Law graduates did not communicate with the *Jury* about legal matters more than other defendants (see Table 5), one may

hypothesize that Law graduates because of their legal training were better at using the available evidence to defend themselves which, in turn, could explain the Law graduate advantage. We test and reject this hypothesis in three ways.

In our first test, we distinguish Law graduates from Parisian Universities from those with a degree from another university outside Paris. The rationale is that the two groups would have similar professional and legal skills and should, therefore, if legal skills were the source of the Law graduate advantage, enjoy the *same* advantage. Appendix C.3.1 presents difference-in-differences estimates of the Law graduate advantage separately for Law graduates from a Parisian university and for Law graduates from another university. We observe that the difference in acquittal rates between Law graduates from a Parisian university and other defendants is larger in front of the *Jury* than in front of the CDLs and is significant at the five-percent level. This is *not* the case for Law graduates from other universities, who do not display any advantage in front of the *Jury*. Appendix C.3.2 moreover shows that the differences in connections observed in the dossiers of defendants also appear only for Law graduates from Parisian universities and not for other Law graduates. This strongly suggests that legal skills cannot explain the Law graduate advantage.

Our second test of legal skills as the source of the Law graduate advantage is based on the idea that Law graduates may have been better than other defendants to present information about their participation in the resistance to the *Jury*. We can test this by augmenting Equation (11) with a triple interaction "Information on participation in resistance ×Jury ×LG". Estimates reported in Appendix Table C.3.3 show that this triple interaction is never significant. Accordingly, Law graduates were not better due to their legal training than other defendants at using information about participation in the resistance in front of the *Jury*.

Our third test is based on the idea that having a political career impairs the accumulation of legal skills. If so, Law graduates who had a longer political career would, therefore, have blunter legal skills, which would, in turn, result in a smaller advantage before the *Jury*. To test this, we augment Equation (11) with a triple interaction "Length of political career ×Jury ×LG". The results reported in Appendix Table C.3.4 show that the Law graduate advantage was larger, rather than smaller, for defendants with longer political careers, measured by the number of years spent as a local representative ("conseiller général"). This, therefore, also suggests that the Law graduate advantage before the *Jury* is unrelated to legal skills.

Overall, the evidence from the various tests consistently militate against the Law graduate advantage being explained by differences in legal skills between Law graduates and other defendants.

6 Conclusion

After a transition to democracy, parts of the old autocratic elite often remain a powerful political force despite attempts to purge those who are compromised by their association with the former regime. An open question is why. We argue that connections within elite groups play an important role. Democratic purges are fundamentally different from those in autocracies in that they follow a structured legal process, and this is what allows connections to play a role. We demonstrate this empirically with evidence from the purge of the French parliamentarians, who had supported the establishment of the Vichy regime, after World War II. The connection mechanism that we identity is of relevance to understanding other similar transition processes such as the Truth and Reconciliation commission in South-Africa, and lustration in post-communist regimes or in post-Pinochet Chile, and adds, more generally, to our understanding of what causes elite persistence.

The analysis has two important implications. First, we stress that in a democratic transition the members of the authority overseeing the purge process are likely to be connected to members of the previous autocratic institutions. As a results, members of the former elite who face the purge can be connected to their judges. Even in the absence of any malicious plan, these connections confer on this subset of the previous elite an advantage in persisting through the transition and in keeping their political influence intact. Second, our results show how distorted political selection in new democracies might be: because of their connections, some elite groups are more likely to persist than others. It thus becomes clear that connections are part of what Acemoglu and Robinson (2006) refer to as *de facto* power. Connections have been proven to be instrumental in distorting laws (Cohen and Malloy, 2014) and biasing political selection within political regimes (Dal Bó et al., 2009). Our analysis shows that connections also matter after major institutional changes and, at least partially, explain elite persistence.

References

Acemoglu, D. and J. A. Robinson (2006). De facto political power and institutional persistence. *American Economic Review 96*(2), 325–330. 2, 3, 31

Acemoglu, D. and J. A. Robinson (2008). Persistence of power, elites, and institutions. *American Economic Review 98*(1), 267–93. 2

Albertelli, S., J. Blanc, and L. Douzou (2019). *La lutte clandestine en France: une histoire de la Reśistance, 1940-1944*. Éditions du Seuil. 4, 5

Alesina, A. and E. La Ferrara (2014). A test of racial bias in capital sentencing. *American Economic Review 104*(11), 3397–3433. 3, 4, 8, 17

Ang, M. and M. Nalepa (2019). Can transitional justice improve the quality of representation in new democracies? *World Politics 71*(4), 631–666. 4

Anwar, S. and H. Fang (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review 96*(1), 127–151. 3, 17

Baruch, M.-O. (2003). *Une poignée de misérables. L'épuration de la société française après la Seconde Guerre mondiale*. Fayard. 5

Bates, G., I. Cinar, and M. Nalepa (2020). Accountability by numbers: A new global transitional justice dataset (1946–2016). *Perspectives on politics 18*(1), 161–184. 2

Berlinski, S., T. Dewan, and K. Dowding (2007). The length of ministerial tenure in the UK 1945-1997. *British Journal of Political Science 37*(2), 245–262. 3

Berlinski, S., T. Dewan, and B. Van Coppenolle (2014). Franchise extension and the b ritish aristocracy. *Legislative Studies Quarterly 39*(4), 531–558. 2

Besley, T., T. Persson, and D. M. Sturm (2010). Political competition, policy and growth: theory and evidence from the US. *Review of Economic Studies 77*(4), 1329–1352. 2

Bindler, A. and R. Hjalmarsson (2018). Path dependency in jury decision making. *Journal of the European Economic Association*. 42

Bueno de Mesquita, B. and A. Smith (2017). Political succession: A model of coups, revolution, purges, and everyday politics. *Journal of Conflict Resolution 61*(4), 707–743. 3

Cohen, L. and C. J. Malloy (2014). Friends in high places. *American Economic Journal: Economic Policy 6*(3), 63–91. 4, 31

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–1451. 10

Dal Bó, E., P. Dal Bó, and J. Snyder (2009). Political dynasties. *The Review of Economic Studies 76*(1), 115–142. 3, 31

Dal Bó, E. and R. Di Tella (2003). Capture by threat. *Journal of Political Economy 111*(5), 1123–1154. 4

Eggers, A., A. Fowler, J. Hainmueller, A. Hall, and J. Snyder (2015). On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *American Journal of Political Science 59*(1), 259–274. 3

Elster, J. (2006). Redemption for wrongdoing: The fate of collaborators after 1945. *Journal of Conflict Resolution 50*(3), 324–338. 5

Fisman, R. (2001). Estimating the value of political connections. *American Economic Review 91*(4), 1095–1102. 4

Fiva, J. and D. M. Smith (2017). Political dynasties and the incumbency advantage in party-centered environments. *American Political Science Review 112*(3), 706–712. 3

Goldring, E. and A. S. Matthews (2021). To purge or not to purge? an individual-level quantitative analysis of elite purges in dictatorships. *British Journal of Political Science*, 1–19. 3

González, F., P. Muñoz, and M. Prem (2021). Large lost in transition? the persistence of dictatorship mayors. *Journal of Development Economics*, 102669. 2, 3

Grossman, G. M. and E. Helpman (2002). *Special Interest Politics*. MIT Press. 10

Higley, J. and M. G. Burton (1989). The elite variable in democratic transitions and breakdowns. *American Sociological Review*, 17–32. 3

Kaminski, M. M. and M. Nalepa (2006). Judging transitional justice: A new criterion for evaluating truth revelation procedures. *Journal of Conflict Resolution 50*(3), 383–408. 2

Knox, D., W. Lowe, and J. Mummolo (2020). Administrative records mask racially biased policing. *American Political Science Review 114*(3), 619–637. 18

Kurrild-Klitgaard, P. (2000). The constitutional economics of autocratic succession. *Public Choice 103*(1/2), 63–84. 3

Lacroix, J., P.-G. Méon, and K. Oosterlinck (2023). A positive effect of political dynasties: the case of france's 1940 enabling act. *Journal of Economics History*. 5, 13

Le Béguec, G. (2003). *La République des avocats*. Armand Colin. 3, 6

Lim, C. S., J. M. Snyder Jr, and D. Strömberg (2015). The judge, the politician, and the press: newspaper coverage and criminal sentencing across electoral systems. *American Economic Journal: Applied Economics 7*(4), 103–35. 4

Linz, J. J., J. J. Linz, A. Stepan, et al. (1996). *Problems of democratic transition and consolidation: Southern Europe, South America, and post-communist Europe*. JHU Press. 2

Marcot, F. (2006). *Dictionnaire historique de la Résistance*. Laffont. 5

Martinez-Bravo, M. (2014). The role of local officials in new democracies: Evidence from indonesia. *American Economic Review 104*(4), 1244–87. 3

Martinez-Bravo, M., P. Mukherjee, and A. Stegmann (2017). The non-democratic roots of elite capture: Evidence from soeharto mayors in indonesia. *Econometrica 85*(6), 1991–2010. 2, 3

Michels, R. (1968). *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. The Free Press. 2, 3

Montagnes, B. P. and S. Wolton (2019). Mass purges: Top-down accountability in autocracy. *American Political Science Review 113*(4), 1045 – 1059. 3

Nalepa, M. (2008). To punish the guilty and protect the innocent: Comparing truth revelation procedures. *Journal of Theoretical Politics 20*(2), 221–245. 2

O'Donnell, G. and P. C. Schmitter (2013). *Transitions from authoritarian rule: Tentative conclusions about uncertain democracies*. JHU Press. 3

Olson, M. C. (1965). *The Logic of Collective Action, Public Goods and the Theory of Groups*. Harvard University Press. 2

Pakulski, J., J. S. Kullberg, and J. Higley (1996). The persistence of postcommunist elites. *Journal of Democracy 7*(2), 133–147. 2

Park, K. H. (2017). Do judges have tastes for discrimination? evidence from criminal courts. *Review of Economics and Statistics 99*(5), 810–823. 4

Paxton, R. O. (1972). *Vichy France: old guard and new order 1940-1944*. Columbia University Press. 4, 5

Querubin, P. (2015). Dynastic incumbency advantage in the philippines. *Quarterly Journal of Political Science 11*(2), 151–181. 3

Shayo, M. and A. Zussman (2011). Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics 126*(3), 1447–1484. 4

Svolik, M. W. (2009). Power sharing and leadership dynamics in authoritarian regimes. *American Journal of Political Science 53*(2), 477–494. 3

Tullock, G. (1987). *Autocracy*. Kluwer. 3

Van Coppenolle, B. (2017). Political dynasties in the uk house of commons: The null effect of narrow electoral selection. *Legislative Studies Quarterly 42*(3), 449–479. 3

Van Coppenolle, B. (2020). How do political elites persist? Political selection, political inequality and empirical history research. *French Politics 18*, 175–180. 2

Voeten, E. (2008). The impartiality of international judges: Evidence from the european court of human rights. *American Political Science Review 102*(4), 417–433. 4

Wieviorka, O. (2001). *Les Orphelins de la République. Destinées des députés et sénateurs français (1940-1945)*. Le Seuil. 5, 13, 18

Wolton, S. (2017). Lobbying, inside and out: How special interest groups influence policy choices. *Available at SSRN 2190685*. 4

A Proofs - Theoretical Section

A.1 Proof - Direct connections

Proof. Consider a given court c. We need to find the values of the fixed cost such that 1) a defendant wants to communicate that $l_{i,c} = \theta_Y$ when that is, in fact, the case and 2) a defendant do not want to claim that $l_{i,c} = \theta_Y$ when $\theta = \theta_N$. Assume that the court believes $l_{i,c} = \theta_Y$ and sets $m_{i,c} = \theta_Y$ in response to a letter with such a claim and consider the game between defendant i and court c. First, suppose that $\theta = \theta_Y$, i.e., there are mitigating circumstances for defendant i. Let the cost of communicating with court c for defendant i be $f_{i,c}$. He will then get $-(\theta_Y - \theta_Y - \gamma)^2 - f_{i,c}$ if he sends the letter $l_{i,c} = \theta_Y$ and $-(\theta_N - \theta_Y - \gamma)^2$ if he does not send a letter. Comparing these payoffs, we see that it is in the interest of defendant i to send the letter $l_{i,c} = \theta_Y$ if

$$f_{i,c} \leq (\theta_Y - \theta_N)(2\gamma + (\theta_Y - \theta_N)) \equiv \bar{f}. \tag{14}$$

Second, suppose that $\theta = \theta_N$, i.e., there are no mitigating circumstances for defendant i. If defendant i has cost $f_{i,c}$, then he will get $-(\theta_Y - \theta_N - \gamma)^2 - f_{i,c}$ if he sends the letter $l_{i,c} = \theta_Y$ and $-(\theta_N - \theta_N - \gamma)^2$ if he does not send a letter. Comparing these payoffs, we see that it is in the interest of defendant i not to lie and send a letter with $l_{i,c} = \theta_Y$ when $\theta = \theta_N$ if

$$f_{i,c} \geq (\theta_Y - \theta_N)(2\gamma - (\theta_Y - \theta_N)) \equiv \underline{f}. \tag{15}$$

Clearly, $\bar{f} > \underline{f}$. Given these strategies, the court will update via Bayes Rule its belief to "mitigating circumstances" if and only if $f_{i,c} \in [\underline{f}, \bar{f}]$. The three cases in the proposition follows immediately from this. □

A.2 Proof - Indirect connections

Proof. This is a standard cheap talk game. Consider court c and suppose that it believes the third party if a letter saying $l_{i,c} = \theta_Y$ is received. There is no reason not send this letter if $\theta = \theta_Y$. Suppose, therefore, that $\theta = \theta_N$. In this case, the third party has an incentive to lie and write in the letter that $l_{i,c} = \theta_Y$. If he does write this, his payoff is $-\beta_{i,c}(\theta_Y - \theta_N)^2 - (1 - \beta_{i,c})(\theta_Y - \theta_N - \gamma)^2$. If he instead writes $l_{i,c} = \theta_N$, then his payoff is $-\beta_{i,c}(\theta_N - \theta_N)^2 - (1 - \beta_{i,c})(\theta_N - \theta_N - \gamma)^2 = (1 - \beta_{i,c})\gamma^2$. Comparing these two payoffs, we find that the third party will not be tempted to write a letter saying $l_{i,c} = \theta_Y$ when $\theta = \theta_N$ if

$$\beta_{i,c} > 1 - \frac{\theta_Y - \theta_N}{2\gamma} \equiv \bar{\beta}. \tag{16}$$

Clearly $\bar{\beta} < 1$. The cut-off $\bar{\beta} > 0$ because we assume that $\gamma > \frac{\theta_Y - \theta_N}{2} \equiv \bar{\gamma}$ for all i. Knowing this, court c will believe a letter of support claiming mitigating circumstances for defendant i coming from a third party with $\beta_{i,c} > \bar{\beta}$ and not otherwise. □

B Baseline results - Robustness checks

B.1 Counterfactual analysis

B.1.1 Counterfactual - Method

We estimated a model where the decision by the *Jury* to clear a defendant was regressed on variables capturing the participation of defendants in civilian and military resistance, which were the criteria officially used by the *Jury*, controlling for the decision by CDLs. We therefore estimated the following regression equation on the sample of decisions of the *Jury*, :

$$Acquitted_{Jury,i} = \alpha + \beta_1 Acquitted_{CDL,i} + \beta_2 CivilianResistance_i + \beta_3 MilitaryResistance_i + \varepsilon_i$$

We then stored the estimators of $\beta_1, \beta_2, \beta_3$: $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ and computed an individual probability to be purged given the information retrieved from defendants' dossiers. The estimated probability given by $Counterfactual_i = \hat{\beta}_1 Acquitted_{CDL,i} + \hat{\beta}_2 CivilianResistance_i + \hat{\beta}_3 MilitaryResistance_i$ provides a counterfactual benchmark based on the official criteria used by the *Jury* against which to compare actual clearance rates. For simplicity, we present our counterfactual at the group-level.

B.1.2 Counterfactual - Comparison with actual acquittal rate

Appendix B.1: Counterfactual versus Actual acquittal rate for non-Law graduates and Law Graduates

	Not Law Graduates	Law graduates
Counterfactual	23.7%	22.4%
Actual	22.8%	30.9%

B.2 Placebo tests and control variables

Appendix B.2: Information in dossier and advantage before the *Jury*

		CDL			Jury			Diff-in-Diff	
		(B.2.1)	(B.2.2)	(B.2.3)	(B.2.4)	(B.2.5)	(B.2.6)	(B.2.7)	
		Treated	Control	Diff=0	Treated	Control	Diff=0	$\Delta\Delta$	p-value
	Treated:	Group	Group	(p-value)	Group	Group	(p-value)		
Control variables	Politics and political mandates								
	Mayor	0.12	0.09	0.31	0.22	0.28	0.14	-0.09**	0.02
	Pres/Vice-Pres or Sec Assembly	0.16	0.10	0.27	0.22	0.25	0.67	0.10**	0.05
	MP of an occupied department	0.09	0.11	0.52	0.20	0.31	0.01***	-0.09**	0.02
	Networks, clubs and religion								
	Jewish MPs	0.17	0.10	0.59	0.67	0.24	0.02**	0.35*	0.09
	Occupations								
	Journalist	0.04	0.11	0.14	0.29	0.25	0.455	0.11*	0.10
	Informational cues								
	Mayor under "Etat Fr"	0.07	0.11	0.16	0.15	0.29	0.00***	-0.09**	0.02
	Arrested by Etat Fr	0.14	0.10	0.56	0.50	0.24	0.01***	0.23**	0.03
	Militarian resistance	0.27	0.06	0.00***	0.68	0.14	0.00***	0.32***	0.00
	Civilian resistance	0.13	0.05	0.01***	0.38	0.08	0.00***	0.22***	0.00
	Continuous variables								
	Age							-0.005**	0.02
Placebo Tests	Politics and political mandates								
	Senator	0.09	0.10	0.95	0.25	0.25	0.92	0.002	0.95
	Rightwing	0.10	0.09	0.42	0.25	0.25	0.93	-0.03	0.49
	Center	0.06	0.11	0.15	0.21	0.26	0.35	0.003	0.95
	MPs elected in Paris	0.05	0.10	0.44	0.25	0.25	0.99	0.05	0.57
	Dynastic Politicians	0.06	0.11	0.29	0.27	0.25	0.70	0.07	0.27
	War experience								
	WWI veteran	0.11	0.09	0.63	0.25	0.25	0.89	-0.008	0.84
	WWII fighter	0.08	0.10	0.78	0.38	0.24	0.15	0.15	0.18
	Networks, clubs and religion								
	Free Masons	0.07	0.10	0.66	0.33	0.25	0.45	0.12	0.30
	Labour Unions	0.03	0.11	0.19	0.19	0.26	0.45	0.01	0.87
	Agr organization	0.13	0.10	0.50	0.26	0.25	0.85	-0.02	0.73
	War Medal	0.11	0.10	0.79	0.24	0.26	0.62	0.03	0.48
	Légion d'Honneur	0.12	0.09	0.21	0.25	0.25	0.88	-0.03	0.46
	Veterans club	0.11	0.10	0.94	0.32	0.25	0.50	0.06	0.60
	Occupations								
	Civil servant	0.08	0.10	0.73	0.32	0.25	0.41	0.10	0.28
	Workers	0.11	0.09	0.90	0.25	0.25	0.99	-0.01	0.91
	Informational cues								
	Excluded by his party	0.12	0.09	0.37	0.31	0.23	0.09*	0.05	0.28
	Signed Bergery motion	0.06	0.11	0.29	0.22	0.26	0.54	0.007	0.89
	Continuous variables								
	National Mandate							0.00	0.76
	Conseiller général							-0.001	0.49
	Study Years							0.003	0.66

Column B.2.1 presents the average acquittal rate of the treated group (defined in the left column) in front of the Comité Départementaux de Libération whereas Column B.2.2 presents the average acquittal rate of the control group (= all individuals not in the treated group) in front of the Comités. Column B.2.3 displays the difference between these two means. Column B.2.4 presents the average acquittal rate of the treated group in front of the Jury whereas Column B.2.5 presents the average acquittal rate of the control group (= all individuals not in the treated group) in front of the Comités. Column 4.6 displays the difference between Column B.2.4 and B.2.5. Column B.2.6 introduces the estimates of Equation 1 without any control for the bias towards each of the subgroup defined in the left column. This estimate is also by construction equal to the difference between Column B.2.6 and B.2.3. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

B.3 Court-level estimates

Appendix B3: Relative bias and courts clearance rate

	(B.3.1)	(B.3.2)
Dependent variable	$Acquit_i$	$Acquit_i$
Samples	CDLs	Jury
LG	-0.0284	0.0712*
	(-0.915)	(1.737)
Constant	-0.0797	0.0636
	(-0.829)	(0.558)
Controls:		
Age and Religion	Yes	Yes
Journalist	Yes	Yes
Political mandates	Yes	Yes
Resistance and collaboration WWII	Yes	Yes
Observations	399	399
Adjusted R-squared	0.088	0.332

Column B.3.1 estimates a bivariate regression estimating the statistical advantage of law graduates before the CDLS and includes all baseline control variables to this estimation. Column B.3.2 estimates a bivariate regression estimating the statistical advantage of law graduates before the Jury and includes all baseline control variables to this estimation. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

B.4 Alternative estimations

Appendix B.4: Alternative speficications

	(B.4.1)	(B.4.2)	(B.4.3)	(B.4.4)	(B.4.5)	(B.4.6)	(B.4.7)	(B.4.8)	(B.4.9)	(B.4.10)	(B.4.11)
Dep Variable	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_i$	$Acquit_i$	$Acquit_i$	$Acquit_i$	$Acquit_i$	$Acquit_i$	$Acquit_i$	$Acquit_i$	$Acquit_i$
Sample	All	All	Jury	Jury	Jury	Jury	Jury	Jury	Jury	Jury	Jury
			All	All	All	Not Acquit CDL	Not Acquit CDL	Not Acquit CDL	Acquit CDL	Acquit CDL	Acquit CDL
Estimation Method	Logit	Probit	OLS	Logit	Probit	OLS	Logit	Probit	OLS	Logit	Probit
Jury × LG	0.977**	0.542**									
	(2.013)	(2.179)									
LG	-0.422	-0.248	0.0834**	0.821**	0.409**	0.111***	1.052***	0.564***	-0.183	-2.218	-0.992
	(-0.975)	(-1.155)	(2.051)	(2.224)	(2.060)	(2.685)	(2.846)	(2.791)	(-1.078)	(-1.081)	(-1.159)
Constant	-4.972***	-2.603***	0.0978	-3.014***	-1.710***	0.0136	-4.173***	-2.380***	1.373***	6.279**	3.499**
	(-3.732)	(-3.819)	(0.949)	(-3.107)	(-3.169)	(0.133)	(-3.819)	(-4.089)	(3.897)	(2.293)	(2.359)
Controls:											
Baseline	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
CDLS decision	NO	NO	YES	YES	YES	NO	NO	NO	NO	NO	NO
Observations	798	798	399	399	399	359	359	359	40	30	30
Adjusted R^2			0.411			0.316			0.107		

Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1. This Table estimates the baseline specification using various estimation methods (Logit and Probit). It also focuses on decisions by the Jury and control for the decision of CDLs various ways. Columns B.4.3 to B.4.5 control for the decision of the CDLs. Columns B.4.6 to B.4.8 estimate the specification within the subsample of cases for which the defendant has not been acquitted by a CDL. Columns B.4.9 to B.4.11 estimate the specification within the subsample of cases for which the defendant has been acquitted by a CDL. This Table presents estimates of Equation 1 when focusing on law graduates and adding variables explaining a difference in sentencing patterns between the Jury and CDL. Individual controls include: Age, Jewishness, Journalist, Mayor, Special Role in the Assembly, Civilian Resistance, Military resistance, Arrested by Etat Francais, Mayor under Etat Francais, Mp of an occupied territory. Each invidual control is also interacted with the Jury dummy variable.

B.5 Controlling for the composition of resistance in each department

Appendix B.5: Controlling for the composition of resistance in each department

	(B.5.1)	(B.5.2)	(B.5.3)	(B.5.4)	(B.5.5)	(B.5.6)	(B.5.7)	(B.5.8)	(B.5.9)	(B.5.10)
Dep variable	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$
Jury	0.133***	0.123***	0.152***	0.114*	0.134***	0.123***	0.118*	-0.0775	-0.110	0.0763
	(5.662)	(4.895)	(7.314)	(1.798)	(5.707)	(4.915)	(1.837)	(-1.030)	(-1.447)	(0.568)
Gaullist/Communist Resistance	0.0738				0.0801			-0.118	-0.122	-0.0839
	(1.039)				(1.081)			(-0.703)	(-0.728)	(-0.479)
Jury × Gaullist/Communist Resistance	0.138				0.152			0.0172	0.0391	0.0483
	(1.277)				(1.372)			(0.0695)	(0.154)	(0.172)
Gaullist/Domestic Resistance		0.0868				0.108		0.446	0.453	0.250
		(0.283)				(0.341)		(0.609)	(0.617)	(0.282)
Jury × Gaullist/Domestic Resistance		0.775				0.849		0.385	0.351	-0.145
		(1.525)				(1.621)		(0.342)	(0.309)	(-0.107)
Foreign/Domestic Resistance			-0.000265		-0.000439	-0.000319	-0.000404*	-0.000237	-0.000214	-0.000230
			(-1.242)		(-1.578)	(-1.223)	(-1.679)	(-0.865)	(-0.756)	(-0.649)
Jury × Foreign/Domestic Resistance			-0.000692***		-0.00102***	-0.00111***	-0.000652***	-0.000607	-0.000718**	-0.000708*
			(-5.750)		(-3.685)	(-3.770)	(-4.462)	(-1.557)	(-2.270)	(-1.862)
Communist/Domestic Resistance				-0.305**			-0.311**	-0.339**	-0.340**	-0.268*
				(-2.519)			(-2.540)	(-2.292)	(-2.298)	(-1.651)
Jury × Communist/Domestic Resistance				0.106			0.100	0.184	0.191	0.155
				(0.602)			(0.568)	(0.825)	(0.864)	(0.696)
LG									-0.0226	-0.0294
									(-0.751)	(-0.950)
Jury × LG									0.110**	0.101**
									(2.414)	(2.183)
Constant	0.0908***	0.0972***	0.101***	0.221***	0.0911***	0.0973***	0.223***	0.169***	0.176***	0.0526
	(5.545)	(5.326)	(6.642)	(4.409)	(5.560)	(5.329)	(4.402)	(3.155)	(3.171)	(0.458)
Proof of resistance								Yes	Yes	Yes
Individual controls										Yes
Observations	798	798	798	798	798	798	798	798	798	798
Adjusted R-squared	0.045	0.043	0.039	0.045	0.046	0.044	0.045	0.251	0.254	0.279

All regressions test the effect of the composition of resistance, and so likely of CDL, on the bias as the ratio of the number of militants belonging to different groups as defined by their certificate of resistance. For example, Gaullist/Communist resistance is defined as the ratio of the number of members of gaullist resistant factions over the number of members of communist resistant factions in a department. Proof of resistance include: Civilian Resistance, Military resistance. Individual controls include the following control variables: Age, Jewishness, Journalist, Mayor, Special Role in the Assembly, Arrested by Etat Francais, Mayor under Etat Francais, In occupied territory. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

B.6 CDLs and the treatment of information on Law graduates

Appendix B.6: The specific treatment of information on Law graduates in front of CDLs

	(B.6.1)	(B.6.2)	(B.6.3)	(B.6.4)	(B.6.5)
Dep Variable	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$
Sample	Decisions CDL	Decisions CDL	Decisions CDL	Decisions CDL	Decisions CDL
LG × Civil Res	0.0831				0.0856
	(1.482)				(1.647)
LG × Military Res		-0.00289			-0.0107
		(-0.0242)			(-0.0904)
LG × Arrested EF			0.171		0.164
			(0.711)		(0.687)
LG × Mayor EF				0.00183	0.00468
				(0.0282)	(0.0753)
Constant	0.0529	-0.0612	0.0971	0.0978	-0.0630
	(0.547)	(-0.656)	(1.056)	(1.069)	(-0.648)
Individual controls	Yes	Yes	Yes	Yes	Yes
Observations	399	399	399	399	399
Adjusted R-squared	0.012	0.079	-0.004	0.003	0.085

Estimations focus on the decisions of CDLs. They assess how the CDLs could have reacted to certain types of information contained in the dossier of the Jury and used by law graduates. They interact the law graduates dummy variable with variable assessing the information in the dossier of defendants. Individual controls include: Age, Jewishness, Journalist, Mayor, Special Role in the Assembly, Mp of an occupied territory. Each invidual control is also interacted with the Jury dummy variable. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

B.7 Persistence: When did the advantage of Law graduates appear?

We proxy the intentions of defendants in two ways. First, we use the information in their dossiers related to participation in the 1945 municipal elections and divide the defendants into two groups: those who ran for election and therefore intended to pursue a political career, and those who did not. Appendix B.7, columns B.7.1 and B.7.2 report separate estimates of equation (11) for the two groups. We observe that the Law graduate advantage is only significant, at the ten-percent level, for the group of defendants who ran in the 1945 municipal election. Moreover, Column B.7.3 reports a specification estimated on the full sample in which equation (11) is augmented with a triple interaction between the Law graduate dummy, the *Jury* dummy, and a dummy coding whether the defendant ran in the election. The coefficient on the triple interaction term is positive and significant at the one-percent level. Its magnitude indicates that the acquittal rate before the *Jury* was 40 percentage points higher for Law graduates who ran for election than for those who did not.

Second, as an alternative proxy for a defendant's intention to seek public office, we leverage a discontinuity caused by a change in the remit of the Jury. Until September 1945, the *Jury* was in charge of judging two types of cases: cases of electoral litigation brought by departmental prefects and cases brought by the defendants themselves. These cases were mainly about eligibility to run for election. An order of 13^{th} of September 1945 expanded the remit of the *Jury* to include the cases of all parliamentarians who had voted in favor of the enabling act or had collaborated with the Vichy regime. Many of whom did not intend to run in an election. We know from the dossiers of the defendants when a case was considered by the *Jury* and can, therefore, distinguish cases considered before and after the change in the remit and create a pre- and a post-reform dummy. In this way, we can use the discontinuity to test if the advantage of Law graduates before the *Jury* was bigger for defendants who wanted to continue their political career (as revealed by an early case related to eligibility for election) than for other defendants. Specifically, we augment equation (11) with interaction terms between the pre- and post-reform dummies and $LG_i \times Jury_c$, respectively, to

allow the Law graduate advantage to differ depending on when the case was heard. Appendix B.7, columns B.7.4 to B.7.6 present the results. The result in column B.7.4 shows that the Law graduate advantage before the *Jury* was around 29 percentage points larger for the defendants tried before the reform when the *Jury* focused on electoral litigation than for those tried after the expansion of its remit.[34] A Wald-test shows that this difference is statistically significant. The specifications in the other two columns are augmented with time polynomials and their interactions with $LG_i \times Jury_c$ and the reform dummy variables and show that that the effect is not driven by time trends in the sentences of the *Jury*.[35] The advantage of Law graduate advantage before the *Jury* therefore facilitated elite persistence, as it materialized specifically when a Law graduate intended to continue his political career.

Appendix B7: Law graduates' advantage appears when it matters: Electoral litigations

	(B.7.1)	(B.7.2)	(B.7.3)	(B.7.4)	(B.7.5)	(B.7.6)
Dep variable	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$
Sample	Ran for elections	Did not run	All	All	All	All
Jury	0.536	0.169	0.165	0.170	0.200*	0.231*
	(1.122)	(1.512)	(1.496)	(1.533)	(1.726)	(1.945)
LG	-0.131	-0.00836	-0.00575			
	(-1.437)	(-0.248)	(-0.170)			
Jury × LG	0.330*	0.0569	0.0524			
	(1.913)	(1.215)	(1.112)			
Jury × LG × Ran for elections			0.404***			
			(2.610)			
Jury × LG X Pre-reform				0.338***	0.281**	0.301**
				(2.985)	(2.349)	(2.462)
Jury × LG × Post-reform				0.0460	0.0661	0.0772
				(0.960)	(1.379)	(1.638)
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial Date					3	3
Pre-reform dummy						Yes
Wald Test				0.016**	0.093*	0.088*
Observations	118	680	798	798	792	792
Adjusted R-squared	0.164	0.303	0.290	0.302	0.336	0.342

Column B.7.1 estimates equation (1) on the subset of defendants who ran for the first post-WWII mayoral elections. Column B.7.2 estimates equation (1) on the subset of defendants who ran for the first post-WWII mayoral elections. Column B.7.3 investigates how the magnitude of the advantage of Law graduates varied with their participation in the first post-WWII elections by adding a triple interaction term (Jury X LG X Ran for elections) and controlling for the interaction Jury X Ran for elections. It contorls for the variables not interacted in the estimation (Jury, LG, and Ran for elections). Columns B.7.4 to B.7.6 estimate equation (1) in a manner akin to a RDD estimates using the cutoff of the September 13rd as a discontinuity. It therefore shows how the reform affected the bias of the Jury towards Law graduates after adding individual controls, a time-polynomial of order 3, and a pre-reform dummy variable. These estimates assess the break in time trend of the advantage of Law graduates due to the reform of the remit of the Jury from mainly electoral litigations to investigations of all cases. Individual controls include: Age, Jewishness, Journalist, Mayor, Special Role in the Assembly, Civilian Resistance, Military resistance, Arrested by by the Vichy regime, Mayor under Etat Francais, parliamentarian of an occupied territory. Each invidual control is interacted with the Jury dummy variable. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

[34] To determine the difference between the two groups we compare the coefficients attached to Law graduates tried before the reform and those tried after the reform, hence: 33.8-4.6.

[35] Part of this difference in sentencing patterns over time might be captured by the comparison with the decision of CDLs (also varying over time). Adding time polynomials however allows to take time varying dynamics of the *Jury* into account as previous research has shown that time-dependence exists in sentencing (Bindler and Hjalmarsson, 2018).

B.8 Persistence and intensive margins

Appendix B.8: Persistence and Jury decision - Intensive margin

Panel A - Independent variable of interest: Acquitted by the Jury

	(B.8.A.1) Years as Mayor	(B.8.A.2) Years in Parliament	(B.8.A.3) Years as Minister
Acquitted_{Jury}	0.271**	0.366***	0.0695**
	(2.006)	(3.113)	(2.032)
Acquitted_{CDL}	-0.0373	0.117	0.0344
	(-0.237)	(0.779)	(0.571)
Constant	1.001***	0.738***	0.0694*
	(4.174)	(3.643)	(1.831)
Observations	399	399	399
Adj R^2	0.155	0.137	0.051
Control variables			
Individual	YES	YES	YES

Panel B - Independent variable of interest: Law graduates

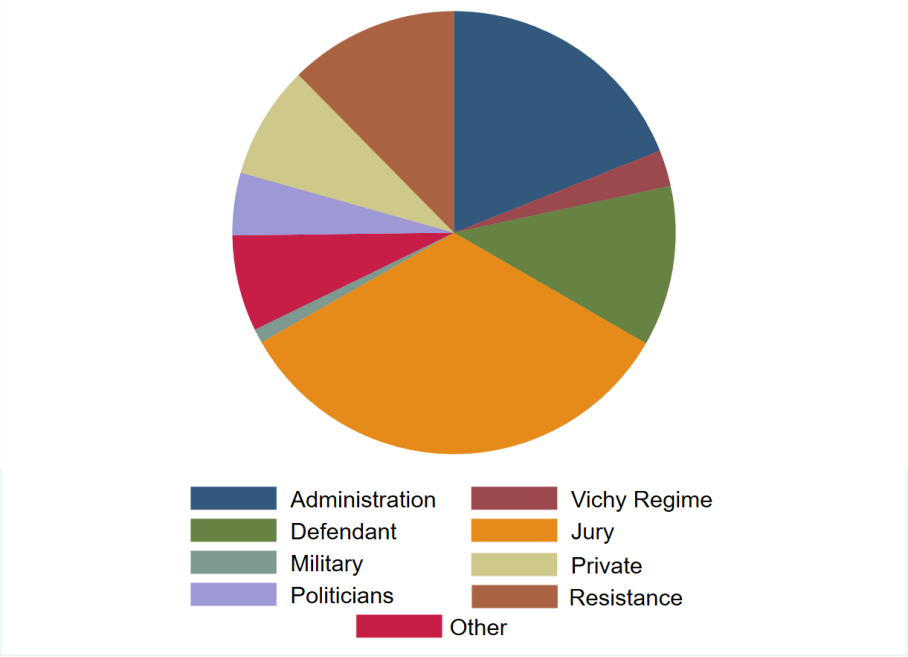
	(B.8.B.1) Years as Mayor	(B.8.B.2) Years in Parliament	(B.8.B.3) Years as Minister
LG	-0.0160	0.181**	0.0341
	(-0.207)	(2.366)	(1.362)
Acquitted_{CDL}	0.0763	0.284*	0.0662
	(0.479)	(1.888)	(1.088)
Constant	1.040***	0.722***	0.0664
	(4.307)	(3.638)	(1.528)
Observations	399	399	399
Adj R^2	0.142	0.115	0.038
Control variables			
Individual	YES	YES	YES

Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Level of observation: Defendant. In Panel A, the main independent variable is a dummy variable equal to 1 if the defendant has been acquitted by the Jury. In Panel B, the main independent variable is a dummy variable equal to one if the defendant was a law graduate. Dependent variables: Years as Mayor is the number of years the defendant has spent as mayor after WWII (Log-transformed). Years in Parliament is the number of years the defendant has spent either as a deputy or as a senator after WWII (Log-transformed). Years as Minister is the number of years the defendant has spent as a minister after WWII (Log-transformed). Individual controls include: Age and religion (Age, Jewishness); Journalist; Political mandates (Mayor, Special Role in the Assembly, parliamentarian of an occupied territory); Resistance and collaboration WWII (Civilian Resistance, Military resistance, Arrested by the Vichy regime, Mayor under Etat Francais).

C Additional evidence on the mechanisms

C.1 Presentation - Structure of the mechanisms dataset

Figure C.1: The origins of all the documents in the dossiers of the defendants



Note: The categories are defined as follows. Administration refers to documents produced by a ministry or a local administration. Vichy Regime refers to documents from the archives of the Vichy regime or from Vichy-related institutions. Defendant refers to documents produced by the defendant himself. Jury refers to documents produced by the *Jury*. Military refers to documents produced by the French army. Private refers to documents produced by an individual in his/her own name without stating an obvious relation to an organization, typically coming from family members or friends of the defendant or from individuals in his constituency. Politicians refers to documents produced by parliamentarians and local politicians. Resistance refers to documents produced by members of resistance networks. Lawyers refers to documents sent by a lawyer using his/her title in the document sent.

C.2 Definition and descriptive stastistics: Mechanisms dataset

Variable	Definition	Min	Max	Mean	s.d
	Structure				
Nb Doc	Number of documents in the dossier	10	170	40.69	26.38
Nb Pages	Number of pages in the dossier	12	384	55.44	45.26
Nb Doc from Jury	Number of Document produced by the Jury	4	50	13.70	6.16
Nb Archival Docs	Number of Archives	0	50	2.79	6.23
Nb Information requests	Number of information requests sent by the Jury	0	4	1.08	0.95
Nb Letters of support	Number of letters of support	0	90	7.81	12.34
Nb Letters of support - in Favor	In favor of acquitting the defendant	0	90	7.41	11.89
Nb Letters of support - Against	Against acquitting the defendant	0	20	0.22	1.31
Nb Letters of support - Neutral	Neutral	0	10	0.18	0.71
Nb Letters of support - Res	from the Resistance	0	39	2.41	4.28
Nb Letters of support - Mil	from the Military	0	11	0.27	1.01
Nb Letters of support - Administration	from an administration	0	41	1.02	2.91
Nb Letters of support - Others	from other type of organizations	0	70	2.66	6.38
Nb Doc - Military resistance	Nb of Doc mentioning participation in military resistance	0	35	0.77	3.11
Nb Doc - Civilian resistance	Nb of Doc mentioning participation in civilian resistance	0	64	8.55	10.63
Nb Doc - Resistant Press	Nb of Doc mentioning participation in resistant press	0	13	0.59	1.79
Nb Doc - Legal Arguments	Nb of Doc mentioning legal arguments	0	27	5.43	3.95
Nb Doc - Political opinion	Nb of Doc mentioning political opinions of the defendant	0	90	10.02	14.34
Nb Doc - Reelection	Nb of Documents mentioning reelection prospects	0	18	1.07	2.37
Nb Doc - Other topic	Nb of Documents mentioning other topics	0	58	6.32	8.65
	Direct connections				
1. Nb Doc from defendant	Nb of Docs sent by the Defendants	0	62	4.64	6.05
1.1 Nb of Doc - from Def without legal content	Nb of Doc sent by the Defendant - No legal content	0	19	1.45	2.23
1.2 Nb of Doc - from Def with legal content	Nb of Doc sent by the Defendant - Legal content	0	55	3.19	4.74
1.3 Nb of pages - from Def without legal content	Nb of Pages sent by the Defendant - No legal content	0	53	1.87	3.81
1.4 Nb of pages - from Def with legal content	Nb of Pages sent by the Defendant - Legal content	0	321	9.68	22.05
2. Nb Informal letters from Defendant	Nb of letters with headings “Cher”	0	6	0.17	0.62
2.1 Nb of Doc - Informal letters without legal content	Nb of letters with headings “Cher” - No legal content	0	3	0.06	0.30
2.2 Nb of Doc - Informal letters with legal content	Nb of letters with headings “Cher” - Legal content	0	6	0.11	0.50
	Indirect connections				
1. Nb Letters of support from Paris	Nb of letters of support by a Parisian sender	0	38	2.07	4.77
2. Nb of Supporters from Paris	Nb of different supporters from Paris	0	14	0.85	1.81
3. Indirect connections via supporters	Average number of documents sent by different supporters (even outside the case)	0	583.83	33.22	78.38
3. Indirect connections via supporters (excluding letters)	Same as above excluding letters	0	341.67	18.29	46.80
5. Nb Informal documents (“Dear” Letters)	Nb of document with headings “Cher”	0	27	1.46	3.02
5.1 Nb Informal documents to Cassin	Addressed to R. Cassin	0	5	0.19	0.56
5.2 Nb Informal documents to Jury	Addressed to the Jury	0	5	0.22	0.61
5.3 Nb Informal documents not to Jury	Not addressed to the Jury	0	26	1.24	2.84

C.3 Advantage not explained by different legal skills

C.3.1 Parisian and non-Parisian Law graduates

Table C3.1: The Law graduate advantage for Parisian vs non-Parisian Law graduates

	Panel A: Control group = All those considered non-treated							
	CDLs			Jury			Diff-in-Diff	
	(C3.1.1)	(C3.1.2)	(C3.1.3)	(C3.1.4)	(C3.1.5)	(C3.1.6)	(C3.1.7)	(C3.1.8)
	Treated	Control	Diff=0	Treated	Control	Diff=0	$\Delta\Delta$	p-value
Considered as treated =	Group	Group	(p-value)	Group	Group	(p-value)		
All Law graduates	0.081	0.11	0.45	0.31	0.23	0.097*	0.11**	0.03
Law graduates (Parisian U)	0.07	0.11	0.37	0.36	0.23	0.04**	0.16**	0.01
Other Law graduates	0.10	0.10	0.96	0.25	0.25	0.94	0.01	0.91
	Panel B: Control group = All non-law graduate							
	CDLs			Jury			Diff-in-Diff	
	Treated	Control	Diff=0	Treated	Control	Diff=0	$\Delta\Delta$	p-value
Considered as treated =	Group	Group	(p-value)	Group	Group	(p-value)		
Law graduates (Parisian U)	0.07	0.11	0.36	0.36	0.23	0.04**	0.17***	0.01
Other Law graduates	0.10	0.11	0.84	0.26	0.23	0.68	0.036	0.60

Note: Panel A provides estimates on the whole sample, i.e., it uses all non-treated defendants are controls. Panel B provides estimates of one subgroup of lawyers compared to non-lawyers, i.e., excluded non-treated Law graduates from the control group. Column C3.1 reports the average acquittal rate of the treated group (defined in the left column) in front of the CDLs whereas column C3.2 reports the average acquittal rate of the control group (= all individuals not in the treated group) in front of the CDLs. Column C3.3 displays the difference between these two means. Column 9.4 presents the average acquittal rate of the treated group in front of the *Jury* whereas column C3.5 presents the average acquittal rate of the control group (= all individuals not in the treated group) in front of the *Jury*. Column C3.6 displays the difference between columns C3.4 and C3.5. Column C3.7 reports the difference-in-differences estimates from Equation (11) (without any control) for the advantage towards each of the subgroups defined in the left column. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

C.3.2 Connections - Parisian and non-Parisian Law graduates

Appendix C.3.2:Documents and parisianism

	Control group = All considered as non-treated					
		Mean			Mean	
	(C3.2.1) LG (Paris U)	(C3.2.2) Control	(C3.2.3) Diff=0	(C3.2.4) LG (not Paris U)	(C3.2.5) Control	(C3.2.6) Diff=0
Difference in:	Group	Group	(p-value)	Group	Group	(p-value)
Nb Letters of support from Paris	3.63	1.80	0.01***	1.69	2.12	0.54
Nb of Supporters from Paris	1.59	0.71	0.00***	0.80	0.85	0.86
Indirect connections via supporters	57.01	29.09	0.01**	36.31	32.77	0.76
Indirect connections via supporters (excl. letters)	32.73	15.78	0.01**	19.49	18.11	0.85
Nb Informal documents ("Dear" Letters)	2.78	1.24	0.00***	1.08	1.52	0.33
Nb Informal documents to Cassin	0.19	0.19	0.98	0.12	0.20	0.34
Nb Informal documents to Jury	0.27	0.21	0.49	0.16	0.23	0.42
Nb Informal documents not to Jury	2.51	1.02	0.00***	0.92	1.29	0.39
	Control group = All non law graduates					
		Mean			Mean	
	LG (Paris U)	Control	Diff=0	LG (not Paris U)	Control	Diff=0
Difference in:	Group	Group	(p-value)	Group	Group	(p-value)
Nb Letters of support from Paris	3.63	1.82	0.01**	1.69	1.82	0.85
Nb of Supporters from Paris	1.59	0.70	0.00***	0.80	0.70	0.65
Indirect connections via supporters	57.01	27.82	0.01***	36.31	27.82	0.46
Indirect connections via supporters (excl. letters)	32.73	15.13	0.01***	19.48	15.13	0.53
Nb Informal documents ("Dear" Letters)	2.78	1.26	0.00***	1.08	1.26	0.63
Nb Informal documents to Cassin	0.19	0.20	0.86	0.12	0.20	0.35
Nb Informal documents to Jury	0.27	0.22	0.58	0.16	0.22	0.49
Nb Informal documents not to Jury	2.51	1.04	0.00***	0.92	1.04	0.73

Upper panel provides estimates on the whole sample. Bottom panel provides estimates of one subgroup of law graduates compared to others. Column C3.2.1 presents the averages in various measures of document contents for Law graduates from a Parisian University; Columns C3.2.4 displays averages for Law graduates from other universities. Columns C3.2.2 and C3.2.5 show the averages for control groups, Columns C3.2.3 and C3.2.6 the difference between averages of the treated group and control group. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

C.3.3 Difference in the treatment of information between CDLs and the Jury - A Law graduate effect?

Appendix C3.3: Difference in the treatment of information between CDLs and the Jury

	(C3.3.1)	(C3.3.2)	(C3.3.3)	(C3.3.4)	(C3.3.5)	(C3.3.6)
Dep Variable	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$	$Acquit_{i,c}$
Sample	All	All	All	All	All	All
Civil Res × Jury	0.141***	0.202***				0.148***
	(3.534)	(5.104)				(3.815)
LG × Civil Res × Jury	0.0652	0.0636				0.0650
	(0.772)	(0.753)				(0.787)
Military Res × Jury	0.260***		0.282***			0.243***
	(3.629)		(4.091)			(3.463)
LG × Military Res × Jury	0.0908		0.0553			0.0628
	(0.555)		(0.358)			(0.390)
Arrested EF × Jury	0.205*			0.223**		0.203**
	(1.943)			(2.156)		(2.045)
LG × Arrested EF X Jury	0.0935			0.117		0.144
	(0.361)			(0.401)		(0.524)
Mayor EF × Jury	-0.0522				-0.0239	-0.00342
	(-1.261)				(-0.487)	(-0.0726)
LG × Mayor EF × Jury	-0.0360				-0.0380	-0.0216
	(-0.385)				(-0.370)	(-0.225)
Constant	0.0630**	0.0529	-0.0612	0.0971	0.0978	-0.0630
	(2.522)	(0.547)	(-0.655)	(1.055)	(1.068)	(-0.647)
Controls		Yes	Yes	Yes	Yes	Yes
Observations	798	798	798	798	798	798
Adjusted R-squared	0.273	0.137	0.226	0.072	0.065	0.284

Columns C3.3.1 to C3.3.6 use the whole set of decisions (both CDL and Jury). They interact a dummy variable for the Jury to variables assessing the information in the dossier and add a triple-interaction of this first term with a law graduate dummy variable to assess the specific reaction of the Jury to this type of information for law graduates. Individual controls include: Age, Jewishness, Journalist, Mayor, Special Role in the Assembly, Mp of an occupied territory. Each invidual control is also interacted with the Jury dummy variable. Robust t-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1

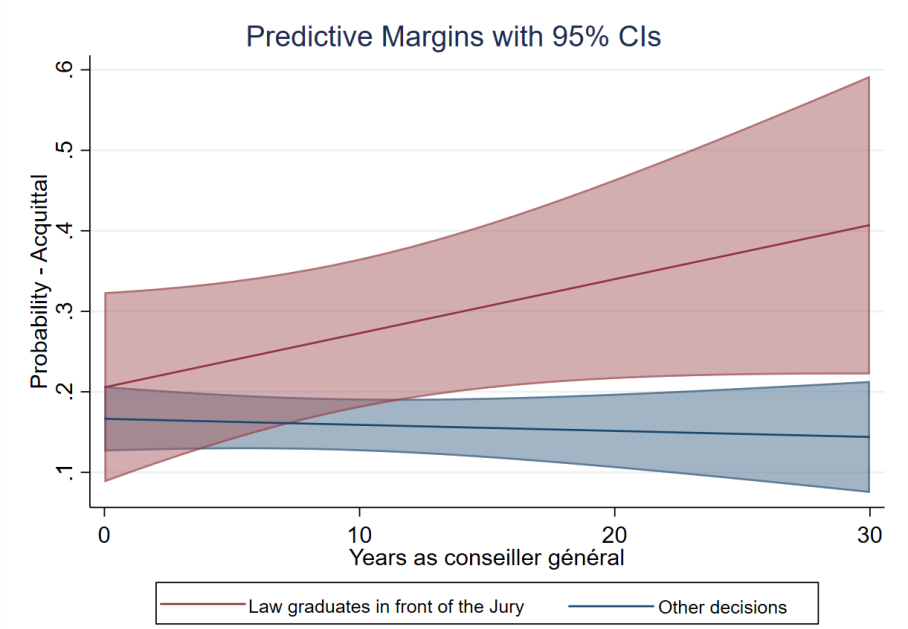
C.3.4 Is the development of legal skills correlating with the bias ?

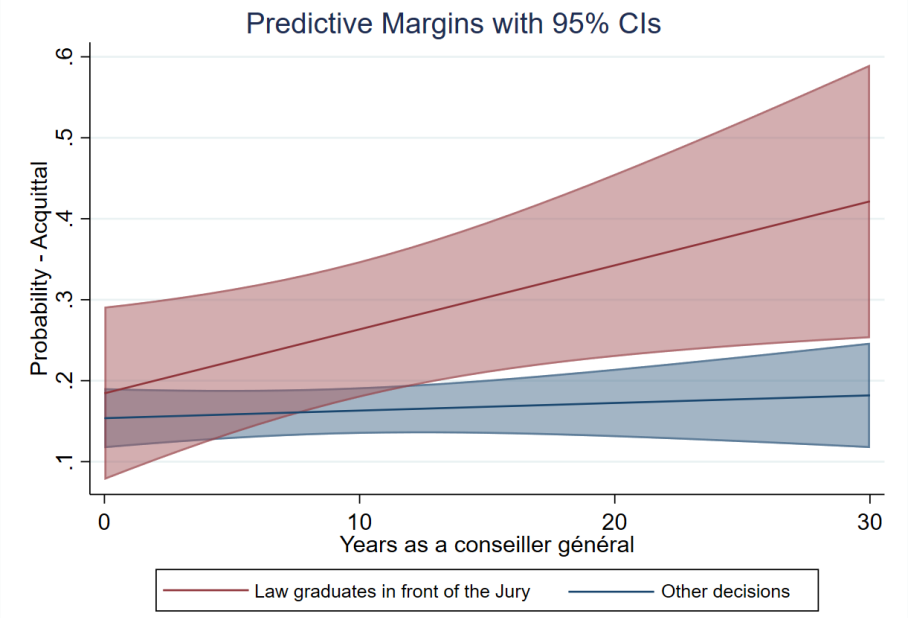
Appendix C3.4: Political career and bias

	(C3.4.1) Diff-in-diff	(C3.4.2) Diff-in-diff
Dependent variable	$Acquit_{i,c}$	$Acquit_{i,c}$
Jury	0.158***	0.185
	(5.344)	(1.610)
LG	-0.0174	-0.0237
	(-0.439)	(-0.579)
LG × Jury	0.0392	0.0307
	(0.604)	(0.511)
LG × Jury × CG	0.00747*	0.00697*
	(1.818)	(1.835)
LG × CG	-0.00115	-0.000988
	(-0.406)	(-0.362)
Constant	0.0938***	-0.0586
	(4.310)	(-0.596)
Full controls		Yes
Observations	798	798
Adjusted R-squared	0.040	0.284

Table C3.4 adds the interaction of the length of the political career with our baseline estimates of the advantage of law graduates in front of the Jury to estimate how renouncing to the development of legal skills affect this baseline effect. Column C3.4 adds full individual control variables. Individual controls include: Age, Jewishness, Journalist, Mayor, Special Role in the Assembly, Mp of an occupied territory. Each invidual control is also interacted with the Jury dummy variable. Robust t-statistics in parentheses:*** p<0.01, ** p<0.05, * p<0.1

Appendix C3.4- Illustrative figure : Marginal effect interaction measure of bias with the length of the political career
(upper panel: without control variables / lower panel: with control variables)





C.4 Correlation Matrix - Measures of indirect connections

Appendix C.4: Correlation matrix - Measures of indirect connections

	1. Nb Letters of support from Paris	2. Nb of Supporters from Paris	3. Indirect connections via supporters	4. Indirect connections via supporters (excl. letters)	5. Nb Informal documents ("Dear" Letters)	5. Nb Informal documents ("Dear" Letters)
1. Nb Letters of support from Paris	\					
2. Nb of Supporters from Paris	0.82	\				
3. Indirect connections via supporters	0.55	0.34	\			
4. Indirect connections via supporters (excl. letters)	0.52	0.32	0.99	\		
5. Nb Informal documents ("Dear" Letters)	0.55	0.52	0.37	0.35	\	
5. Nb Informal documents ("Dear" Letters)	0.54	0.52	0.37	0.35	0.98	\

D Descriptive statistics - Different datasets

D.1 Description of Defendants characteristics

Appendix D.1: Description - Defendants' characteristics

Variable	Mean	s.d	Min	Max
Law graduates	0.28	0.45	0	1
Dependent variable				
Acquittal in front of Jury	0.25	0.43	0	1
Acquittal in front of CDL	0.10	0.30	0	1
Politics and political mandates				
Senator	0.35	0.48	0	1
Rightwing	0.51	0.50	0	1
Center	0.21	0.41	0	1
Mayor	0.47	0.50	0	1
Pres/Vice-Pres or Sec Assembly	0.08	0.27	0	1
MPs elected in Paris	0.05	0.22	0	1
MPs of an occupied department	0.52	0.50	0	1
Dynastic politicians	0.16	0.37	0	1
War experience				
WWI Veteran	0.51	0.50	0	1
WWII fighter	0.06	0.24	0	1
Networks, clubs and religion				
Free Mason	0.04	0.19	0	1
Jewish MPs	0.02	0.12	0	1
Labour unions	0.08	0.27	0	1
Agricultural organizations	0.10	0.29	0	1
Légion d'Honneur	0.38	0.48	0	1
War Medal	0.38	0.49	0	1
Veterans club	0.05	0.21	0	1
Occupation				
Civil Servant	0.06	0.24	0	1
Workers	0.07	0.26	0	1
Journalist	0.12	0.33	0	1
Informational cues				
Excluded by his party	0.27	0.44	0	1
Signed Bergery motion	0.13	0.33	0	1
Mayor under "Etat Fr"	0.27	0.45	0	1
Arrested by Etat Fr	0.06	0.23	0	1
Militarian resistance	0.20	0.40	0	1
Civilian resistance	0.58	0.49	0	1
Continuous variables				
Age	60.25	10.43	34	84
National Mandate	11.19	8.07	1	38
Conseiller Général	8.63	10.67	0	42
Study Years	3.42	3.14	0	8

D.2 Mechanisms: documents dataset

Appendix D.2: Summary statistics - Documents dataset

Documents	Total	Min per dossier	Max per dossier	Mean per Dossier	s.d
All type of document	17589	1	170	40.62	26.9
		Produced by			
Jury	5882	0	50	13.58	6.13
Defendant	2061	0	62	4.76	6.27
Administrations	3335	0	48	7.70	6.00
Politicians	802	0	26	1.85	3.26
Resistant Organizations	2176	0	45	5.03	5.6
Journal Officiel	441	0	3	1.01	0.38
Press	251	0	24	0.58	1.96
Private	1449	0	70	3.34	7.28
Military	181	0	12	0.42	1.34
Intelligence Services	162	0	9	0.37	0.96
Lawyer	115	0	10	0.27	1.03
Court	131	0	6	0.30	0.96
Labor Unions	38	0	4	0.09	0.39
Veterans Association	41	0	7	0.09	0.56
Collaborationist administration	466	0	19	1.07	2.39
Others	58	0	5	0.12	0.54
		Type of documents			
Letters of support	3385	0	90	7.54	12.09