

Are Citizens Tougher on Politicians Than Other Professions? Evidence from Survey Experiments in the United States and Canada

Abstract

How do citizens hold elected officials accountable? We approach this crucial question by exploring how citizens' judgments of politicians' personal conduct differ from their judgments of identical conduct attributed to individuals from other professions. We designed three original survey experiments covering a range of negative and positive conduct—tax evasion, alleged sexual harassment, and positive work ethic—and administered these to broadly representative samples of Americans and Canadians. Contrary to the commonly held view that politicians are held in unusually low regard, we find that citizens generally do not judge politicians more harshly than our two benchmark professions: lawyers and family doctors. More specifically, citizens tend not to reserve harsher punishments for politicians, nor are they notably less willing to give politicians the benefit of the doubt in the absence of definitive information about their conduct. The findings shed new light on longstanding debates about the quality of democratic accountability.

Keywords: Politicians; Public Opinion; Accountability; Survey Experiment

The ability to hold elected officials accountable for their performance and conduct in office is widely and justifiably seen as a leading feature of modern democratic government (Burke 1774; Mansbridge 2003; Pitkin 1967). In an ideal model, citizens are empowered to remove corrupt, immoral, or even simply underperforming individuals from office and to replace them with representatives they believe can do better. The normative appeal of this accountability mechanism has, in turn, spurred empirical research that examines how citizens hold elected officials accountable. This naturally raises the question: by what standard should we judge citizens' judgments? One prominent approach explores whether citizens' judge incumbent performance on the basis of prevailing conditions or events that are not entirely (or even partly) within the incumbent's control—on the grounds that such misplaced judgments are evidence that citizens do not live up to the ideal of democratic accountability (e.g., Achen and Bartels 2017; for a review, see Healy and Malhotra 2013). In this study, we contribute a fresh perspective to this literature by investigating how citizens hold elected officials accountable for their personal conduct. To do so, we develop and test new theoretical benchmarks for judging such judgments—namely, by comparing how citizens judge the personal conduct of politicians to how they judge identical conduct attributed to individuals from other types of professions.

We begin with the premise that being an elected official is a job into which you are hired and from which you can be fired. To be sure, the job of being a politician is not identical to other professions. For instance, few other jobs are filled by an election. Nevertheless, being a politician bears important similarities to other lines of work: it is a position of responsibility; it frequently involves the supervision of subordinates; it can be time-consuming and stressful; and it involves actions that are not always directed observed by others. We are agnostic about whether citizens should

judge conduct by politicians and other professionals in an identical manner. Some might argue for the same standards regardless of profession—perhaps on the grounds that systematically unfavourable assumptions about politicians or unfair levels of scrutiny risk dissuading high-quality potential candidates from pursuing a career in politics. Others might argue that politicians—who are in positions of significant public trust—should be held to higher standards of conduct than other professions and that, by choosing a public life, politicians have accepted the greater scrutiny that comes with this role. Regardless of one’s normative position, however, we contend that comparing citizens’ judgments of personal conduct across professions sheds important new light on debates about the state of accountability in modern representative government and deepens our understanding of how citizens judge their elected representatives.

We proceed as follows. First, we summarize current research about how citizens judge politicians, focusing on work that suggests that citizens tend to engage in mistaken or unfair reasoning about their elected officials. Second, we develop new theoretical standards for assessing how citizens judge politicians’ personal conduct. Third, we present our overall empirical strategy: three pre-registered survey experiments administered to broadly representative pools of US and Canadian citizens. The experiments cover a range of personal behaviour—tax evasion, alleged sexual harassment, and positive work ethic—and allow us to directly compare citizens’ judgments of politicians and other types of professionals in both negative and positive contexts. Next, we present the findings of each experiment in succession. Contrary to a commonly held view that the public holds politicians in unusually low regard, we find that citizens generally do not judge politicians more harshly than our two benchmark professions: lawyers and family doctors. Rather, the picture that emerges across the three experiments is one in which citizens endorse broadly

similar punishments for misconduct across professions. Furthermore, we show that—in the absence of conclusive information about the target individual’s actions—citizens make broadly similar assumptions about conduct across professions. We conclude by discussing the normative implications of these findings for democratic governance and by highlighting possible directions for future research.

Judging citizens’ judgments

Political scientists have devoted considerable attention to how citizens judge the performance of political incumbents—and rightly so given its centrality to our notions of effective representative government (e.g., Fiorina 1981; Key 1966). Unfortunately, there are several reasons to suspect that citizens may not live up to the ideal of careful and reflective judges of their electoral officials.

First, prior research suggests that citizens, at times, judge incumbents on the basis of seemingly irrelevant events—such as rewarding politicians when the local college football team wins (Healy, Malhotra, and Mo 2010) or punishing politicians for natural disasters over which they presumably have little control (Achen and Bartels 2017; Bodet et al. 2016). These findings have, in turn, given rise to vigorous scholarly debate and replication efforts (Busby and Druckman 2018, Busby et al. 2017; Fowler and Hall 2018; Fowler and Montagnes 2015; Graham et al. 2020).

Second, citizens judge politicians unfairly on the basis of their medical history. For instance, citizens are significantly less supportive of political candidates who suffer from depression than from certain physical illnesses. Moreover, the electoral penalty for depression is not a function of the amount of work missed by the politician on account of the illness (Loewen and Rheault 2021).

Similarly, Magni and Reynolds (2021) show that citizens punish political candidates with HIV—which they attribute in part to public prejudice and moral judgments about candidates’ personal responsibility.

Third, public opinion data suggest that politicians as a profession are held in unusually low regard. Gallup, for example, regularly asks American citizens how they rate the honesty and ethical standards of different fields. Since 1976, when participants were first asked about Members of Congress, the results have been consistent: the profession has been repeatedly rated among the least honest. Data from other well-established democracies—such as Canada (Insights West 2017) and the United Kingdom (Ipsos Mori 2017; see also Ram 2006)—show a similar pattern. The consistency of these profession ratings, over time and across countries, point to a stereotype of the corrupt and lazy politician—compared to, for example, the honest and diligent doctor. Such a stereotype would seem detached from the real-world lives of many politicians whose work can be demanding and stressful (Weinberg and Cooper 2003; Weinberg 2011).

In short, the portrait of citizens’ judgments painted by these prior findings is not flattering. Nevertheless, we argue that the picture remains incomplete. Research that assesses the quality of citizen judgment by estimating the extent to which such judgments are influenced by irrelevant events does not—by design—speak directly to judgments of politicians’ personal conduct. While prior studies highlight how citizens’ judgments of politicians’ personal health conditions are subject to prejudice or bias, it remains unclear whether such prejudice is reserved for politicians in particular or whether it is more universal in application. Furthermore, survey data about citizens’ views of generic professions in the abstract fails to take the cross-profession comparison a step

deeper: to examine whether specific instances of personal conduct are subject to different standards when citizens judge the behaviour of different types of professionals. Given the importance of citizens' evaluations of politicians' personal character and competence (e.g., Bittner 2011), we argue that it is important to investigate how—if at all—citizens' judgments of specific instances of conduct and misconduct differ across professions.

Comparing judgments across professions

We begin by proposing two standards for comparing citizens' judgments of personal conduct across different professions: a punishment standard and a “benefit of the doubt” standard.

A punishment standard. Do citizens punish individuals from different professions differently -- even if they engage in the exact same misconduct? If citizens are tougher on politicians, as their generic ratings of their profession would suggest, then we would expect the proposed punishment to be greater for politicians than other professionals.

A benefit of the doubt standard. When information about personal conduct is uncertain, do the assumptions and inferences citizens make differ across professions? Often by necessity, citizens routinely make judgments on the basis of incomplete or uncertain information. If citizens are tougher on politicians, then we would expect that the assumptions they make about the conduct of a politician will be less favorable than the assumptions they make about the conduct of individuals from other professions.

Next, we turn to the question of which professions would serve as informative benchmarks of politicians' conduct. In selecting these benchmark professions, our main goal was to choose two professions that citizens rate—a the more abstract, generic level—more favourably than politicians: one profession that is somewhat more favourably rated and another that is much more favourably rated. The aim in doing so is to establish the scope of any differences in citizens' judgments across professions. For example, are citizens tougher on politicians compared only to highly-rated professions, or are they tougher relative to middle-rated professions as well? By the same token, if we find that citizens are not tougher on politicians compared even to highly-rated professions, then we might conclude they are not any tougher on politicians than most—perhaps nearly all—other professions.

For a profession rated only somewhat more favourably than politicians, we selected lawyers and, for a profession rated substantially more favourably than politicians, we selected doctors. This ordering of professions from low to high ranking—politicians, then lawyers, then doctors—holds both in the US and Canada. For example, in Gallup's December 2020 poll, only 8% of Americans said they believed that Members of Congress had either high or very high honesty and ethical standards. The comparable figures for lawyers and doctors were 21% and 77%, respectively (Gallup 2020). Similarly, in a 2017 poll, only 24% of Canadians reported having a positive view of politicians—compared to 50% for lawyers and 89% for doctors (Insights West 2017).

Empirical approach and data

We test these two standards using these two professional benchmarks across a range of scenarios using three pre-registered between-subject survey experiments.¹ Each experiment concerns the conduct of a hypothetical individual who is identified as either a politician, a lawyer, or a family doctor.²

We fielded the three experiments to representative pools of US and Canadian citizens.³ Our aim in conducting the experiments separately in each country is to shed light on the cross-national generalizability of results. While the US and Canada tend to share broadly similar political values (Dalton 2019), American political discourse tends to be more oriented around candidate personalities than its northern neighbor (e.g., Dalton, McAllister and Wattenberg 2000).

The US survey (N=1,490) was fielded from December 22, 2020 to January 8, 2021. Participants were recruited using a commercial sample provider. The sample was constructed using quotas for

¹ The experiments were pre-registered with EGAP. A copy of the anonymized pre-analysis plan can be found in the online appendix. The study was conducted under REB Protocol XXXX at [university name redacted at review stage to preserve author anonymity].

We deviate from the pre-analysis plan in one respect. In both the US and Canadian surveys, we included two standard attention checks. In the PAP, we indicated that participants who failed one or more attention checks would not be included in the analysis. We do this for the US sample in the manuscript -- presenting the results for participants who passed both attention checks. For the Canadian sample, we deviate from the PAP by presenting the results for participants who passed the first attention check but not necessarily the second. This was done so as to maintain a comparable sample size in both countries. In the appendix (Tables A5-A7), we report the pre-registered Canadian analyses -- that is, including only participants that passed both attention checks. The two sets of Canadians results generate very similar estimates. Where they differ substantively, we note these in the main text.

² In the experiments, we used the term “family doctor” rather than simply “doctor” so that it was clear to participants in both countries (including Canadians who completed the survey in French) that the individual was a medical doctor who provides primary care to patients—rather than, for example, a university professor with a PhD. Experiments involving hypothetical individuals are very common. Reviewing highly ranked journals in political science, McDonald (2020: 269) found that the use of survey experiments making use of hypothetical, fictitious or unfamiliar politicians were twice as numerous experiments using known politicians. Compared to real-world stimuli, hypothetical individuals enhance experimenter control and ensure that participants were not ‘pre-treated’ with prior real-world information about the individual.

³ In both surveys, the order of the three experiments was randomized.

gender, age, racial identity, Hispanic identity, and region based on national census figures. In the analysis that follows, the data are weighted using survey weights constructed using an iterative proportional fitting (IPF) algorithm based on national census figures for XXXX. The Canadian survey (N=1,481) was conducted from May 21 to June 11, 2021, again using a commercial sample provider. Participants could choose to complete the survey in either English or French. National quotas based on gender, age, province, and--within Quebec--French-speaking participants were used in sample construction. In the analysis, we used again IPF weights—in this instance, based on national population parameters for XXXXX.

Experiment 1: Conviction for tax evasion

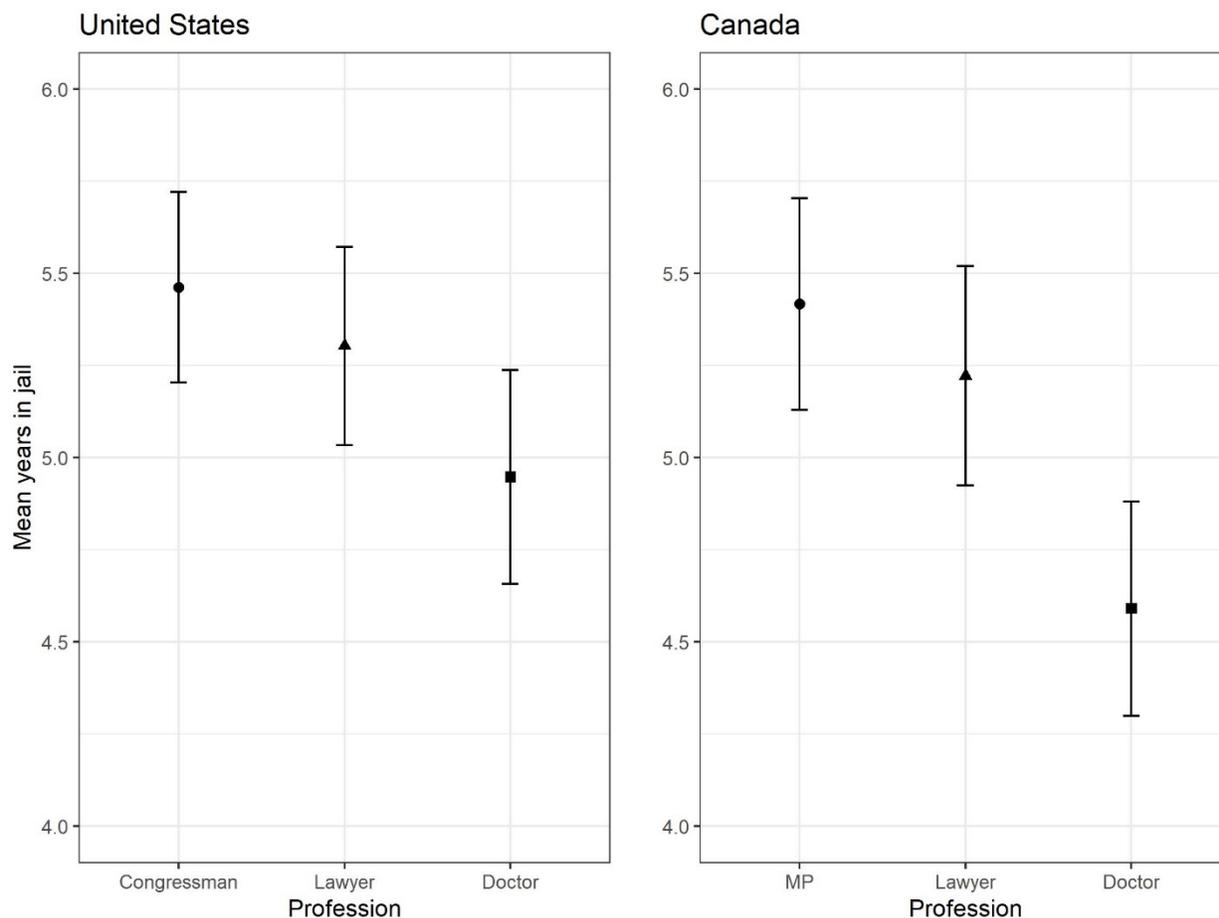
In the first experiment, participants were asked to “imagine the following scenario”:

Recently, Joseph Leonard, a [congressman (US), Member of Parliament (Canada) / lawyer / family doctor], was convicted of tax evasion. Leonard was found to have hidden income from a private business that he owned. In all, he failed to pay \$520,000 in income taxes over the past six years.

Participants were randomly assigned to one of three conditions that varied the individual’s profession: either a politician, a lawyer, or a family doctor. The wording of the *politician* condition was tailored to each country: in the US, the individual was identified as a “congressman” while the individual was identified as a “Member of Parliament” in Canada. After reading the vignette, participants were asked “How many years in prison (from 0 years to 10 years) do you think he should be sentenced to?” Thus, Experiment 1 provides a first test of the *punishment standard*. Given the high degree of certainty regarding the misconduct (by design, Leonard is “convicted”), if participants favour sentencing the politician to a longer prison term than the lawyer and the family doctor, then we have reason to believe that—under these circumstances—citizens judge politicians more harshly than other professionals.

To estimate the treatment effects, we regressed participants' preferred prison sentence (in years) on binary indicators of the *lawyer* and *doctor* conditions (the *politician* condition serves as the omitted baseline). OLS regressions with robust standard errors—estimated separately for the US and Canadian samples—are reported in the appendix. From these results, Figure 1 reports the estimated means for each condition, along with their associated 95% confidence intervals. The results make clear that citizens do *not* choose a stiffer sentence for the politician compared to the lawyer. In both countries, the average jail term for the politician is approximately five and a half years; the difference between the politician and the lawyer is not statistically significant.

Figure 1. Preferred jail sentence by profession and country



Note: This figure reports the estimated means for each condition, along with their associated 95% confidence intervals. See Table A1 for the full regression results.

Figure 1, does, however show that citizens typically wanted the family doctor punished less than the politician. This is the case in both countries. In the US, citizens wanted the doctor to receive, on average, 0.8 fewer years in prison ($p < .01$); in Canada, citizens wanted the doctor to spend 0.5 fewer years in prison ($p < .001$) compared to the politician. Notably, however, the estimated difference in the order of 5-8% of available scale range. Thus, while we do find evidence, on the basis of the *punishment standard*, that citizens are tougher on politicians than family doctors, the substantive effect size is modest.

Experiment 2: Sexual misconduct

In the second experiment, participants were again asked to imagine a scenario—this time involving an individual who was accused of sexually harassing a female employee. The experiment employed a 3 X 2 factorial design. As with Experiment 1, participants were randomly assigned to one of three conditions that varied the target individual’s profession. In addition, the amount of evidence supporting the allegation was randomly assigned: half of participants read only about the allegation itself (the *allegation only* condition); the other half read about how an internal investigation found “a great deal of evidence to support the allegation” (the *supporting evidence* condition). The full vignette text reads as follow:

Denis Martin, a [congressman (US), Member of Parliament (Canada) / lawyer / family doctor], was recently accused by a current female employee of repeatedly making unwelcome, sexually suggestive comments to her at work. He has denied the allegation.

[BLANK / An internal investigation has found a great deal of evidence to support the allegation.]

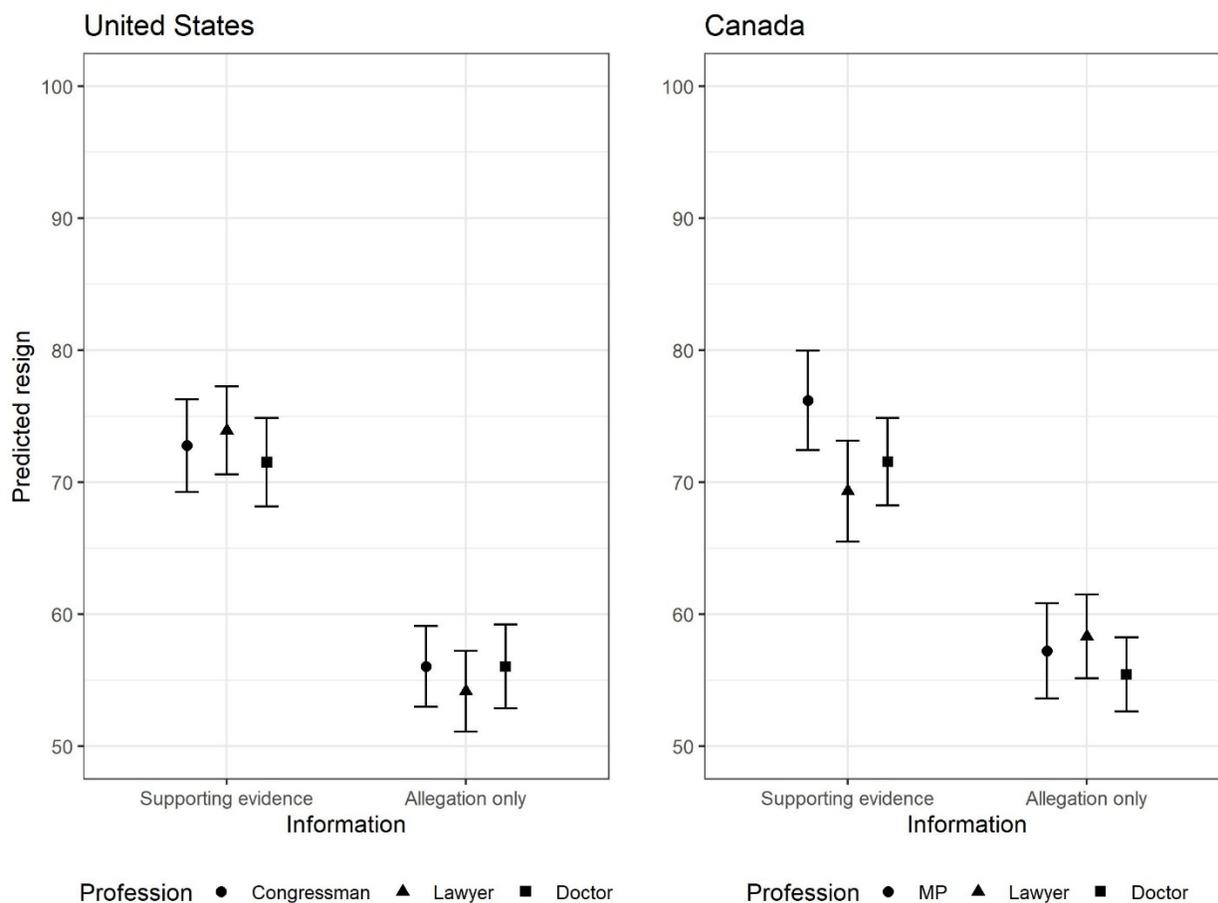
Following the vignette, participants were asked two questions: one in which they indicated how much they believed the allegation (0 = entirely false; 100 = entirely true); and a second in which they indicated how strongly they believed the individual should stay on in their job or resign (0 = should definitely stay on; 100 = should definitely resign).

Experiment 2 allows us to assess both the *punishment standard* and the *benefit of the doubt standard*. To test the *punishment standard*, we compare citizens’ judgments across professions within the *supporting evidence* condition—where (like in Experiment 1) participants have a high degree of certainty that the individual committed the misconduct. If citizens punish politicians more than other professions, then we would expect participants to signal a stronger preference for

the politician to resign than for the lawyer or family doctor to resign. To test the *benefit of the doubt standard*, we compare the effect of the *allegation only* condition—relative to the *supporting evidence* condition—across professions. If citizens are less likely to give politicians the benefit of the doubt, then we would expect that the effect of *allegation only* to be more negative for the lawyer and the family doctor than for the politician. Put another way, the expectation here is that judgments under certainty and uncertainty would differ less for the politician than for the other two professionals.

We estimated treatment effects by regressing participants' preference that the individual resign on binary indicators of the *lawyer* and *doctor* conditions (the *politician* condition serves again as the omitted baseline), a binary indicator of the *allegation only* condition (the *supporting evidence* condition serves as the omitted baseline), an interaction between *lawyer* indicator and the *allegation only* indicator, and an interaction between the *doctor* indicator and *allegation only* indicator. The models—estimated separately for each country using OLS regression with robust standard errors—are reported in Table A2 of the appendix. Figure 2 below presents the estimated means for each condition, and their 95% confidence intervals, as derived from these regression models.

Figure 2. Preference for resignation by profession and country



Note: This figure reports the estimated means for each condition, along with their associated 95% confidence intervals. See Table A2 for the full regression results.

We begin by examining the *punishment standard*: that is, judgments across professions within the *supporting evidence* condition. We find no evidence that Americans are more likely to want to the congressman to resign—compared to the lawyer or family doctor. We do, however, find evidence that Canadians seek to punish politicians more than lawyers for this kind of misconduct: compared to the Member of Parliament, participants felt approximately 6.9 points less strongly—on the original 0 to 100 scale—that the lawyer should resign ($p < .05$). Citizens also felt less strongly

that the family doctor should resign, but the difference only approaches conventional levels of statistical significance ($p = .12$).

Turning to the *benefit of the doubt standard*, an initial point to observe is that the supportive results of the internal investigation are highly consequential for citizens' judgments. Compared to the *supporting evidence* condition, participants in both countries felt approximately 15-20 points less strongly (again, on the original 0 to 100 scale) that the individual should resign. The key test for the *benefit of the doubt standard*, however, is whether this reduction is less steep for politicians than for lawyers or doctors. To restate, because we expect citizens to be less likely to give politicians the benefit of the doubt, we expect the *gap* between the *supporting evidence* and *allegations only* conditions to be *smaller* for politicians than for the other professions. The results in Figure 2 show no evidence of this. In neither country is the effect of the *allegation only* treatment more negative for the lawyer than the politician or for the family doctor than the politician. In fact, in the Canadian sample, the results seem to point in the opposite direction: we find that the negative effect of *allegation only* is roughly 7 points smaller for the lawyer than the Member of Parliament ($p < .05$) and 5 points smaller for the doctor than the Member of Parliament ($p < .05$).

To provide a supplementary test of the benefit of the doubt standard, we also estimated a model in which the *belief-in-allegations* outcome (0 = entirely false; 100 = entirely true) was regressed on binary indicators of the *lawyer* and *family doctor* conditions (politician again serves as the omitted reference category). Here examine only the response of those exposed to the *allegation-only* condition (i.e., the condition in which participants did not have clear supporting evidence). The regression results are reported in Tables A3 and AX. The results provide no evidence that citizens

found that allegations made against the lawyer and the family doctor to be any less credible than identical allegations made against the politician.

In short, we find substantively small and inconsistent evidence that citizens want politicians punished more for sexual misconduct than other types of professionals. In addition, the experiment suggests that—in absence of supportive evidence—citizens do not make more negative assumptions about a politician’s conduct than the conduct of a lawyer or family doctor. When it comes to instances of sexual harassment, citizens’ presumptions are quite similar across the three professions.

Experiment 3: Dedication to the job

In the third experiment, we turn to exploring citizens’ judgments of *favourable* conduct across the three professions. Like the second experiment, Experiment 3 employs a vignette-based, 3 X 2 factorial design. Here the hypothetical individual’s profession is again randomly assigned as above. In addition, we manipulate the level of positive information about the individual’s conduct on the job. Half of participants read that the individual was hardworking and committed to his work (the *hardworking* condition). The other half of participants read an otherwise identical vignette that omitted this information (the *no information* condition). The wording of the vignette is as follows:

Vincent Forrest is a [congressman (US), Member of Parliament (Canada) / lawyer / family doctor].

[BLANK / He works about 60 hours per week. He is fully dedicated to his job and is always available to respond to urgent requests outside of typical work hours.]

After reading this vignette, participants were asked to indicate how much money they thought the individual should earn in a year. Using a slider, participants could select any whole number from \$0 to \$500,000. The range of the scale was designed to allow participants a large window in which to assess the value of the individual's work.⁴

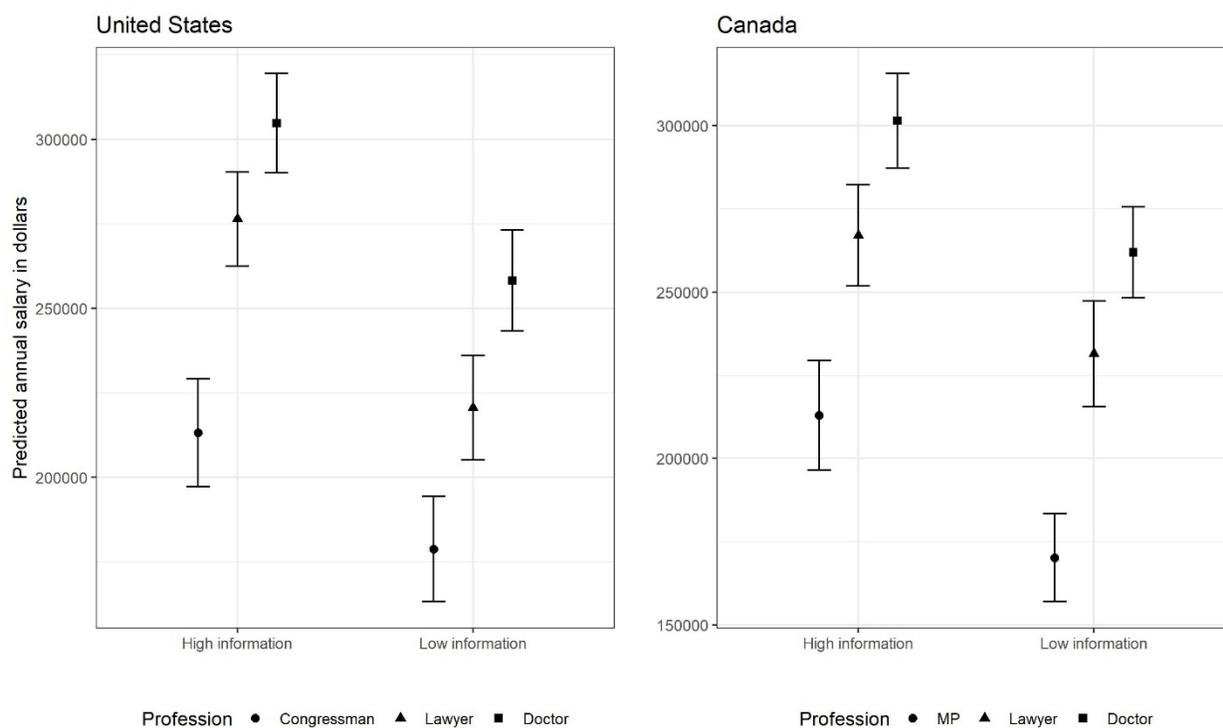
Experiment 3 is designed to evaluate only the *benefit of the doubt* standard. To do so, we compare the effect of the *no information* condition—relative to the *hardworking* condition—across professions. If citizens are less likely to give politicians the benefit of the doubt—in this instance, less likely to assume favourable conduct in the absence of strong evidence—then we would expect that the effect *hardworking* condition would lead to a stronger positive effect for the politician than for the lawyer or the doctor. In other words, we would expect assumptions of good conduct by the lawyer and doctor to yield comparatively small salary differences between the *hardworking* and *no information* conditions for these two professions. By contrast, if citizens do not make these same favourable assumptions about the politician, then we would expect a sharper drop in preferred salary when participants are not explicitly told about the politician's hard work. Thus, the experiment is designed to assess the effect of performance uncertainty against *profession-specific* baselines: that is, the average preferred salary for that profession when citizens knew that the individual was hardworking.

To test the *benefit of the doubt* standard, we replicate the modeling strategy used in Experiment 2. Preferred annual salary (in dollars) are regressed on binary indicators of the *lawyer* and *doctor* conditions (the *politician* being the omitted baseline), a binary indicator of the *no information*

⁴ For reference, a member of the US House of Representatives earns about USD 175,000 and a Member of the Canadian Parliament makes about CAD 185,800 (Congressional Research Service 2021; House of Commons 2018).

condition (the *hardworking* condition serves as the omitted baseline), an interaction between *lawyer* indicator and the *no information* indicator, and an interaction between the *doctor* indicator and *no information* indicator. The OLS models, computed separately for the US and Canadian samples, are reported in Table A4 of the appendix. The model predictions, along with their 95% confidence intervals, are presented in Figure 3.

Figure 3. Preferred annual salary by profession and country



Note: This figure reports the point estimates, and associated 95% confidence intervals, from model 2 (United States) and model 4 (Canada) of Table A4.

Perhaps not surprisingly, citizens believe that politicians should earn less than lawyers and family doctors and the differences are substantively large. It is noteworthy that, even in the *hardworking* condition where participants are provided with clear information that the individual worked hard and was committed to the job, there remain striking disparities in preferred salary: in both the US and Canada, participants suggested that a hardworking politician should earn approximately \$90,000 less than a hardworking family doctor.

While notable, however, these salary differences—on their own—do not constitute a test of whether citizens are tougher on politicians than other professionals. Such differences could simply arise from how citizens value the work of being a politician, a lawyer, or a doctor—or from their assumptions about the real-world salaries of these professions. Rather, the critical test is whether the drop in salary when moving from the *hardworking condition* to the *no information* condition is more negative for politicians than for lawyers and doctors—this would suggest that politicians are less likely to receive the benefit of the doubt. Figure 3 does indeed show a drop in salary for all three professions: when citizens are *not* told about the individual's work ethic, they tend to think that he should earn less than when they are told about his commitment to the job. Crucially, however, this drop in salary does not detectably differ between the politician and the lawyer or between the politician and the family doctor in either the US or Canadian samples. Thus, the results of Experiment 3 lend further support to the conclusion suggested in Experiment 2: namely that, contrary to expectations, citizens are *not* tougher on politicians according to the *benefit of the doubt* standard.

Discussion

Across the three pre-registered experiments and two national samples, the picture that emerges is one in which politicians are not judged significantly more harshly than individuals from other professions. First, we find very little support for the idea that citizens punish the personal misconduct of a politician more than they punish identical conduct by a lawyer and a family doctor—either in the case of tax evasion or sexual harassment. Where we do find differences, these effects are inconsistent across countries and substantively small. On average, Canadians and Americans preferred a slightly longer jail sentence for a politician convicted of tax evasion—but only compared to a family doctor and, even then, the substantive difference is small given the high-status professional benchmark of the doctor. Additionally, when presented with scenarios in which an internal investigation found strong evidence of sexual harassment, Canadians did feel more strongly that the politician should resign than the lawyer—but the same pattern is not apparent in the US sample. Second, when presented with a higher degree of uncertainty about conduct, we find no evidence that citizens make more negative assumptions about a politician than a lawyer or a doctor. This result holds in both cases of misconduct (sexual harassment) and good conduct (dedication to the job) and in both the American and Canadian samples.

In short, when it comes to politicians, our results suggest that citizens are not substantially more punitive—nor are they less likely to give politicians the “benefit of the doubt.” This conclusion is especially striking in light of the considerable survey evidence that citizens, in the abstract, tend to rate politicians as a profession very unfavourably. In particular, the absence of strong differences between judgments of politicians (often among the lowest-rated professions) and doctors (often among the highest-rated professions) is notable and would suggest the absence of such differences between politicians and most other professions.

Our experiments do, however, have limitations that point to potential avenues for future research. First, the individuals featured in all of our experimental stimuli were male. This design choice was made to maximize our statistical ability to detect differences across professions, but it also reflects the fact that men remain significantly over-represented among the ranks of US and Canadian elected officials. It remains an open and important question whether our findings generalize to women politicians. Previous research shows that, under certain conditions, citizens hold men and women politicians to different standards and, notably in the present context, tend to see women politicians as more honest than their male counterparts (Barnes and Beaulieu 2018; Eggers, Vivyan and Wagner 2018; Schwarz and Coppock 2021). Future work should assess how well our experimental findings replicate for women in politics, law, and medicine—and whether women politicians are treated differently when compared to their male counterparts.

Second, by design, we do not specify the partisanship of the individuals in the experimental stimuli. To be sure, partisanship (in particular, a citizen's in-partisans and out-partisans) influences how citizens judge personal conduct among politicians (Walter and Redlawsk 2019). We omitted partisanship because the presence of party labels are uncommon in routine exposure to lawyers and doctors. Including party labels in these conditions would have weakened the realism of those conditions. Moreover, providing a party label for the politician but not for the lawyer or the doctor would have meant confounding profession and party in the experiment. Future work may wish to explore this point further by employing another baseline profession for which party information is more readily accessible—such as judges chosen through partisan elections (Canes-Wrone et al. 2012).

Accountability is crucial for democratic governance—but existing work on the topic does not provide any clear insights as to whether citizens’ evaluations of politicians and politician conduct are unusual when compared to their evaluations of similar conduct by other professionals. The experiments presented here fill this gap and shed light on how citizens attribute punishment—and reward—to politicians, family doctors and lawyers. The results suggest that—in large part—politicians are not held to a different accountability standard when compared to other professions in cases of misconduct. We also do not find evidence to suggest that generally low levels of support for politicians result in a-priori bias toward politicians: none of our experimental conditions suggested that politicians were less likely to be given the benefit of the doubt.

Where we observe clear differences, however, is in the realm of *reward*. Politicians are allocated much lower levels of rewards for their labour—even when participants are informed that politicians work hard. These differences may reflect perceptions of actual salaries, or they may reflect generally lower levels of appreciation for the services provided by politicians compared to family doctors and lawyers. The fact that the public thinks politicians should not receive salaries comparable to lawyers and doctors matters not just in terms of the rewards we think public servants should receive, but also because comparatively lower politician salaries likely affect who select into the profession—which in turn affects the diversity of democratic representation.

Across experiments we find that information provision is highly consequential for citizens to make accountability judgments. Moving from low to high information settings changes citizens’ evaluations and judgments for all professions, both in the case of misconduct and in the case of

reward for hard work. This emphasizes the importance of transparency and information provision—such as the role of the media—for the accountability mechanism to work.

References

Achen, C., Bartels, L., Achen, C. H., & Bartels, L. M. (2017). *Democracy for realists*. Princeton University Press.

Barnes, T. D., & Beaulieu, E. (2019). Women politicians, institutions, and perceptions of corruption. *Comparative Political Studies*, 52(1), 134-167.

Bittner, A. (2011). *Platform or personality?: the role of party leaders in elections*. OUP Oxford.

Bodet, M. A., Thomas, M., & Tessier, C. (2016). Come hell or high water: An investigation of the effects of a natural disaster on a local election. *Electoral Studies*, 43, 85-94.

Burke, E. (1774). *Mr. Edmund Burke's Speeches at His Arrival at Bristol: And at the Conclusion of the Poll*. J. Wilkie.

Busby, E. C., & Druckman, J. N. (2018). Football and Public Opinion: A Partial Replication and Extension. *Journal of Experimental Political Science*, 5(1), 4-10.

Busby, E. C., Druckman, J. N., & Fredendall, A. (2017). The political relevance of irrelevant events. *The Journal of Politics*, 79(1), 346-350.

Canes-Wrone, Brandice, Tom S. Clark, and Jee-Kwang Park. 2012. "Judicial Independence and Retention Elections." *Journal of Law, Economics, and Organization* 28(2): 211–34.

Congressional Research Service. 2021. Salaries of Members of Congress: Recent Actions and Historical Tables. Online: <https://fas.org/sgp/crs/misc/97-1011.pdf> [accessed June 2nd 2021]

Dalton, Russell J. 2019. "Political Culture and Values." In *The United States and Canada: How Two Democracies Differ and Why It Matters*, ed. Paul J. Quirk. New York: Oxford University Press.

Dalton, R. J., McAllister, I., & Wattenberg, M. P. (2000). The consequences of partisan dealignment. *Parties without partisans: Political change in advanced industrial democracies*, 37-63.

Dionne, E. J. Jr. 2004. *Why Americans Hate Politics*. New York: Simon and Schuster

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2020). How Affective Polarization Shapes Americans' Political Beliefs: A Study of Response to the COVID-19 Pandemic. *Journal of Experimental Political Science*, 1-12.

Eggers, A. C., Vivyan, N., & Wagner, M. (2018). Corruption, accountability, and gender: Do female politicians face higher standards in public life?. *The Journal of Politics*, 80(1), 321-326.

Enos, Ryan D., Mark Hill and Austin M. Stragne. 2017. Volunteer Laboratories for Social Science. Working paper. <https://scholar.harvard.edu/files/renos/files/enoshillstrange.pdf>

Fiorina, M. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

Fowler, A., & Hall, A. B. (2018). Do shark attacks influence presidential elections? Reassessing a prominent finding on voter competence. *The Journal of Politics*, 80(4), 1423-1437.

Fowler, A., & Montagnes, B. P. (2015). College football, elections, and false-positive results in observational research. *Proceedings of the National Academy of Sciences*, 112(45), 13800-13804.

Gallup. 2018. Honesty/Ethics in Professions. <http://news.gallup.com/poll/1654/honesty-ethics-professions.aspx>

Gies, Lieve. 2011. "Starts Behaving Badly". *Feminist Media Studies* 11(3): 347-361.

Graham, M. H., Huber, G. A., Malhotra, N., & Mo, C. H. (2020). Irrelevant Events and Voting Behavior: Replications Using Principles from Open Science.

Green, J., Edgerton, J., Naftel, D., Shoub, K., & Cranmer, S. J. (2020). Elusive consensus: Polarization in elite communication on the COVID-19 pandemic. *Science Advances*, 6(28), eabc2717.

Healy, A., & Malhotra, N. (2013). Retrospective voting reconsidered. *Annual Review of Political Science*, 16, 285-306.

Healy, A. J., Malhotra, N., & Mo, C. H. (2010). Irrelevant events affect voters' evaluations of government performance. *Proceedings of the National Academy of Sciences*, 107(29), 12804-12809.

House of Commons. 2018. Members' Allowances and Services Manual. Online: <https://www.ourcommons.ca/Content/MAS/mas-e.pdf> [accessed June 2nd 2021]

Human, R., & Bondevik, K. M. (2011). Fighting stigma with openness. *Bull World Health Organ*, 89, 862-863.

Insights West. 2017. Nurses, Doctors and Scientists Are Canada's Most Respected Professionals. <https://insightswest.com/news/nurses-doctors-and-scientists-are-canadas-most-respected-professionals/>

Ipsos MORI. 2017. Politicians Remain the Least Trusted Profession in Britain. <https://www.ipsos.com/ipsos-mori/en-uk/politicians-remain-least-trusted-profession-britain>

Key, V.O. 1966. *The Responsible Electorate*. New York: Vintage

Krook, M. L. (2018). Violence against women in politics: A rising global trend. *Politics & Gender*, 14(4), 673-675.

Loewen, P. J., & Rheault, L. (2021). Voters punish politicians with depression. *British Journal of Political Science*, 51(1), 427-436.

Magni, G., & Reynolds, A. (2021). The Persistence of Prejudice: Voters Strongly Penalize Candidates with HIV. *Political Behavior*, 1-20.

Mansbridge, J. (2003). Rethinking representation. *American political science review*, 97(4), 515-528.

McDonald, J. (2020). Avoiding the hypothetical: why "Mirror experiments" are an essential part of survey research.

NL Times. Economic Affairs Minister Steps Aside for 3 Months Over Burnout. <https://nltimes.nl/2021/05/24/economic-affairs-minister-steps-aside-3-months-burnout> [accessed June 2nd 2021]

Nye, Joseph Jr, Philip D. Zelikow and David C. King, ed. 1997. *Why People Don't Trust Government*. Cambridge: Harvard University Press.

Pitkin, H. F. (1967). *The concept of representation*. University of California Press.

Ram, V. (2006). Public attitudes to politics, politicians and parliament. *Parliament Affairs*, 59(1), 188-197.

Rheault, L., Rayment, E., & Musulan, A. (2019). Politicians in the line of fire: Incivility and the treatment of women on social media. *Research & Politics*, 6(1), 2053168018816228.

Schwarz, S. and Coppock, A. 2021. "What Have We Learned About Gender From Candidate Choice Experiments? A Meta-analysis of 67 Factorial Survey Experiments", *Journal of Politics*.

Walter, A. S., & Redlawsk, D. P. (2019). Voters' partisan responses to politicians' immoral behavior. *Political Psychology*, 40(5), 1075-1097.

Weinberg, A. (Ed.). (2011). *The psychology of politicians*. Cambridge University Press.

Weinberg, A., & Cooper, C. L. (2003). Stress among national politicians elected to Parliament for the first time. *Stress and Health: Journal of the International Society for the Investigation of Stress*, 19(2), 111-117.

Online Appendix

Table A1: Effects of profession on preferred jail sentence (0-10 year scale)

	(1)	(2)
	<i>US</i>	<i>Canada</i>
Profession: Lawyer	-0.159 (0.190)	-0.195 (0.211)
Profession: Doctor	-0.515** (0.198)	-0.827*** (0.208)
Intercept	5.462*** (0.132)	5.417*** (0.146)
N	1490	1481
R ²	0.005	0.013

Note: OLS regressions with robust standard errors and weights applied.
The reference category for professions is the politician.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Effects of profession and information on preference for resignation (0 to 100 scale)

	<i>United States</i>	<i>Canada</i>
	(1)	(2)
Profession: Lawyer	1.150 (2.472)	-6.871* (2.735)
Profession: Doctor	-1.280 (2.477)	-4.654* (2.556)
Information	-16.726*** (2.375)	-18.974*** (2.661)
Profession: Lawyer × Information	-3.049 (3.316)	7.971* (3.676)
Profession: Doctor × Information	1.268 (3.345)	2.868 (3.465)
Intercept	72.785*** (1.790)	76.198*** (1.917)
N	1490	1481
R ²	0.104	0.090

Note: OLS regressions with robust standard errors and weights applied.

The reference category for professions is the politician. The reference category for the “Allegation only” condition is the “Supporting evidence” condition.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: Effects of profession on the believability of the allegation (0 to 100 scale)

	<i>United States</i>	<i>Canada</i>
	(1)	(2)
Profession: Lawyer	-1.191	2.312
	(1.754)	(1.860)
Profession: Doctor	0.554	2.184
	(1.760)	(1.786)
Intercept	59.294***	60.271***
	(1.200)	(1.384)
N	756	738
R ²	0.001	0.003

Note: OLS regressions with robust standard errors and weights applied.

The reference category for professions is the politician.

* $p < 0.5$, ** $p < 0.01$, *** $p < 0.001$

Table A4. Effects of profession and information on preferred salary (\$0 to \$500,000 scale)

	<i>United States</i>	<i>Canada</i>
	(1)	(2)
Profession: Lawyer	63219.451*** (10788.100)	53993.093*** (11438.333)
Profession: Doctor	91573.154*** (11067.292)	88484.229*** (11082.152)
Information	-34403.543** (11369.215)	-42927.893*** (10749.667)
Profession: Lawyer × Information	-21396.130 (15544.177)	7370.250 (15563.783)
Profession: Doctor × Information	-12165.963 (15605.907)	3368.643 (14716.832)
Intercept	213213.555*** (8132.449)	213058.009*** (8381.760)
N	1490	1481
R ²	0.113	0.131

Note: OLS regressions with robust standard errors and weights applied. The reference category for professions is the politician. The reference category for the “No information” condition is the “Hardworking” condition.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5. Robustness check: Effects of profession on preferred jail sentence (Canadian sample—excluding participants who failed one or more attention checks)

	(2)
	<i>Canada</i>
Profession: Lawyer	-0.233 (0.284)
Profession: Doctor	-0.720** (0.269)
Intercept	5.227*** (0.191)
N	879
R ²	0.009

Note: OLS regressions with robust standard errors and weights applied.

The reference category for professions is the politician.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6. Robustness check: Effects of profession and information on preference for resignation (Canadian sample—excluding participants who failed one or more attention checks)

	<i>Canada</i> (2)
Profession: Lawyer	-6.476 (3.738)
Profession: Doctor	-4.651 (3.344)
Information	-21.968*** (3.535)
Profession: Lawyer × Information	8.893 (4.951)
Profession: Doctor × Information	1.342 (4.509)
Intercept	77.146*** (2.573)
N	879
R ²	0.122

Note: OLS regressions with robust standard errors and weights applied.

The reference category for professions is the politician. The reference category for the “Allegation only” condition is the “Supporting evidence” condition.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7. Robustness check: Effects of profession and information on preferred salary (Canadian sample—excluding participants who failed one or more attention checks)

	<i>Canada</i> (2)
Profession: Lawyer	56374.667*** (13333.872)
Profession: Doctor	99457.724*** (14093.240)
Information	-49612.365*** (12432.890)
Profession: Lawyer × Information	5254.696 (18089.990)
Profession: Doctor × Information	4308.863 (18291.182)
Intercept	203963.619*** (9694.129)
N	879
R ²	0.172

Note: OLS regressions with robust standard errors and weights applied. The reference category for professions is the politician. The reference category for the “No information” condition is the “Hardworking” condition.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$