

# Electoral Accountability in Multiparty Contests<sup>1</sup>

Andrew C. Eggers<sup>†</sup>

Nick Vivyan<sup>‡</sup>

Markus Wagner<sup>§</sup>

## Abstract

We develop a theoretical framework and conduct a survey experiment to assess electoral accountability in multiparty plurality contests. Compared to two-party elections, multiparty elections offer voters additional channels with which to respond to incumbent performance, with subtle but potentially important implications for electoral accountability. We provide conditions under which more non-Duvergerian competition (i.e. greater support for uncompetitive candidates) is associated with stronger punishment for incumbent misconduct; our analysis suggests that these conditions are generally unlikely to be met. We then conduct a survey experiment on British voters to estimate parameters from our model. Our findings suggest that the non-Duvergerian nature of many plurality elections in the UK weakens electoral accountability in that setting, and the same may be true in Canada, India and other contexts where multiple parties win support in plurality contests.

---

<sup>1</sup>First version: June, 2014. This version: December 5, 2014.

<sup>†</sup>Nuffield College and Department of Politics and International Relations, University of Oxford, UK. email: [andrew.eggers@nuffield.ox.ac.uk](mailto:andrew.eggers@nuffield.ox.ac.uk)

<sup>‡</sup>School of Government and International Affairs, Durham University, UK. email: [nick.vivyan@durham.ac.uk](mailto:nick.vivyan@durham.ac.uk)

<sup>§</sup>Department of Methods in the Social Sciences, University of Vienna, Austria. email: [markus.wagner@univie.ac.at](mailto:markus.wagner@univie.ac.at)

# 1 Introduction

Almost without exception, the literature on electoral accountability presumes that voters face a binary choice: typically, incumbent vs. challenger. In reality, however, voters essentially always face more than two choices, and even in plurality systems (notwithstanding the classic Duvergerian claim) many voters support candidates that finish in third place or lower (Raymond, 2013; Singer, 2013). While the binary approach has the advantage of simplicity, it overlooks patterns of voter behavior that may be consequential for understanding electoral accountability, particularly in plurality systems like the UK, India, or Canada where uncompetitive candidates receive substantial support. For example, from the incumbent’s perspective it is more damaging if one of her supporters defects to the main challenger than if the same supporter defects to an uncompetitive candidate, yet these actions would be treated the same in a binary approach that focuses only on the incumbent’s electoral support.

In this paper, we consider how electoral accountability operates in plurality elections when voters can choose among multiple parties. In particular, we ask whether poor incumbent performance is punished more or less heavily in a non-Duvergerian context compared to a simple two-candidate race. This question is particularly relevant because of the substantial proportion of voters who in fact support uncompetitive candidates in plurality elections in many systems. For example, in the 2010 general election in the UK, support for candidates finishing third or lower was 24% on average across constituencies; in both Canada in 2011 and India in 2014 the comparable number was 21%.<sup>2</sup> Because the literature focuses on a binary type of accountability, it is unclear whether we should expect voters to be more or less responsive to incumbent behavior in systems where the baseline support for uncompetitive candidates is higher. Of course, if voters who support uncompetitive candidates will do so regardless of incumbent performance, then we should expect electoral accountability to work poorly in highly non-Duvergerian systems. On the other hand, if support for uncompetitive candidates is sufficiently responsive to incumbent performance then accountability may be better in a non-Duvergerian setting than in a strictly two-party setting. That is, if poor incumbent performance would cause many vot-

---

<sup>2</sup>Non-Duvergerian outcomes have been explained by coordination failure, expressive voting, and voter time horizons that extend beyond a single election (Cox, 1997; Ferree, Powell and Scheiner, 2014).

ers to switch from a non-competitive candidate to the main challenger, then we might expect harsher overall punishments for incumbents in a multiparty contest than in a two-party contest. After all, the only voters who can punish incumbent misconduct in a two-party setting are those voters who would ordinarily support the incumbent and can withdraw that support in response to misconduct; in a multiparty contest, by contrast, punishment can take place not just through the withdrawal of support for the incumbent but also through the shift in support from an uncompetitive candidate to a competitive challenger.

Because prior theoretical work has focused on a two-party context, we begin by developing a theoretical framework for analyzing non-Duvergerian accountability. In our simple decision-theoretic model, an incumbent candidate faces one main challenger and one uncompetitive candidate; we assume that voters act on the basis of a combination of expressive and instrumental motives. We highlight two ways in which voters can respond to an episode of incumbent misconduct: “Strong” punishment involves a voter switching in response to misconduct from the incumbent to the main challenger, while “Weak” punishment involves a voter switching in response to misconduct from the incumbent to the uncompetitive candidate or from the uncompetitive candidate to the main challenger. We can then express the relationship between the punishment the incumbent receives and the baseline support for the uncompetitive candidate in terms of the proportion of voters engaging in each kind of punishment.

We then carry out a survey experiment that allows us to estimate parameters of our model with a representative sample of British voters. We present respondents with a series of hypothetical election scenarios in which the respondent is told that only one challenger has a realistic chance of defeating the incumbent, and is also told either that the incumbent has been commended for “diligent and ethical service by a Westminster watchdog” or that the incumbent has “inappropriately claimed over £10,000 in expenses”. The incumbent’s conduct and the party of the incumbent and main challenger are jointly randomized, with many respondents seeing the same match-up with both a good and bad incumbent. The design allows us to estimate, for different party match-ups, what proportion of voters respond to incumbent misconduct by engaging in either “Strong” or “Weak” punishment; given estimates of vote choice in the “good incumbent” case, we can then estimate how accountability varies with support for uncompetitive

candidates.

Our principal finding is that punishment is less severe as competition becomes less Duvergerian, holding fixed the margin between the incumbent and the challenger. We highlight a tradeoff that takes place as contests become more non-Duvergerian: on one hand there are more voters who would support uncompetitive candidates in the absence of incumbent misconduct, and who could punish a bad incumbent by switching to the main challenger; on the other hand there are fewer voters who would support the incumbent in the absence of incumbent misconduct, and who could punish a bad incumbent by switching to the main challenger or an uncompetitive candidate. Our finding is that, in our survey, the rate of punishment in the first group (supporters of uncompetitive candidates) is small compared to the rate of punishment in the second group (supporters of the incumbent), with the net result that overall punishment goes down as competition becomes less Duvergerian. The parameters on which we base this analysis are of course specific to the context we study, but we provide conditions under which the conclusions would apply to less stark episodes of misconduct; more broadly, our framework and research design can be applied to study non-Duvergerian accountability in other settings.

Although we maintain a substantive focus on multiparty contests in this paper, it should be noted that our theoretical framework and overall research design apply equally well to the question of how the level of abstention (or, conversely, voter turnout) affects electoral accountability. We see the same accountability tradeoff in elections with high levels of abstention as in elections with high levels of support for uncompetitive candidates: abstainers, like supporters of uncompetitive candidates, can punish incumbent misconduct by supporting the main challenger, but it is unclear *ex ante* whether this channel of accountability is strong enough that the net effect of higher baseline abstention on punishment is positive or negative. Our experiment was not designed to separately analyze the role of abstention in accountability, and most of our analysis collapses abstention into support for uncompetitive candidates, but future work could build on our framework to separately study the tendency of non-voters to throw their support behind the challenger in the event of incumbent misconduct.

## 2 Strong and weak punishments for misconduct in a multiparty setting

In this section we provide a formal framework for thinking about how voters respond to misconduct when their options are broader than simply voting for the incumbent or challenger. In particular, we consider a contest in which voters with both expressive and instrumental goals choose between an incumbent, a main challenger, and an uncompetitive candidate; the focus is not on defining optimal behavior of voters or candidates or characterizing equilibrium in strategic interactions, but rather on the simpler task of characterizing the ways in which voter behavior mechanically responds to changes in the attractiveness or perceived quality of the incumbent, given the presence of multiple candidates.

### 2.1 Voter preferences

A voter faces three options: an incumbent  $I$ , a main challenger  $A$ , and an uncompetitive candidate  $B$ . The utility a voter expects from voting for a given candidate  $x$  reflects both expressive and instrumental motives. That is, the voter gets expressive utility from the act of voting for a given candidate, and this utility varies across candidates according to her tastes; she also gets instrumental utility from casting the decisive vote that results in the election of one candidate as opposed to another, and this utility also varies across candidates according to her (policy) preferences. The voter's decision is then based on the expected utility of each possible vote, given the probability of various election outcomes. More formally, the expected utility of voting for candidate  $x$  can be expressed

$$V(x) = U_{\mathcal{E}}(x) + \sum_{x' \neq x} \pi(x, x') \left( U_{\mathcal{I}}(x) - U_{\mathcal{I}}(x') \right) \quad (1)$$

where  $U_{\mathcal{E}}(x)$  measures the expressive (act-dependent) value of voting for candidate  $x$ ,  $\pi(x, x')$  measures the probability of casting the pivotal vote between candidate  $x$  and  $x'$ , and  $U_{\mathcal{I}}(x)$  measures the instrumental (outcome-dependent) benefit of electing candidate  $x$  as opposed to  $x'$ .

We make two simplifying assumptions to streamline the analysis. First, we assume that the

instrumental benefit of electing a given candidate is linearly related to the expressive benefit of voting for that candidate, i.e.

$$kU_{\mathcal{E}}(x) = U_{\mathcal{I}}(x) \equiv U(x).$$

Second, we assume that candidate  $B$  has no chance of winning, which means that  $\pi(I, B) = \pi(A, B) = 0$ , but  $\pi(I, A) = \pi > 0$ . Given these assumptions, we have that

$$V(I) = (1 + k\pi)U(I) - k\pi U(A) \tag{2}$$

$$V(A) = (1 + k\pi)U(A) - k\pi U(I) \tag{3}$$

$$V(B) = U(B). \tag{4}$$

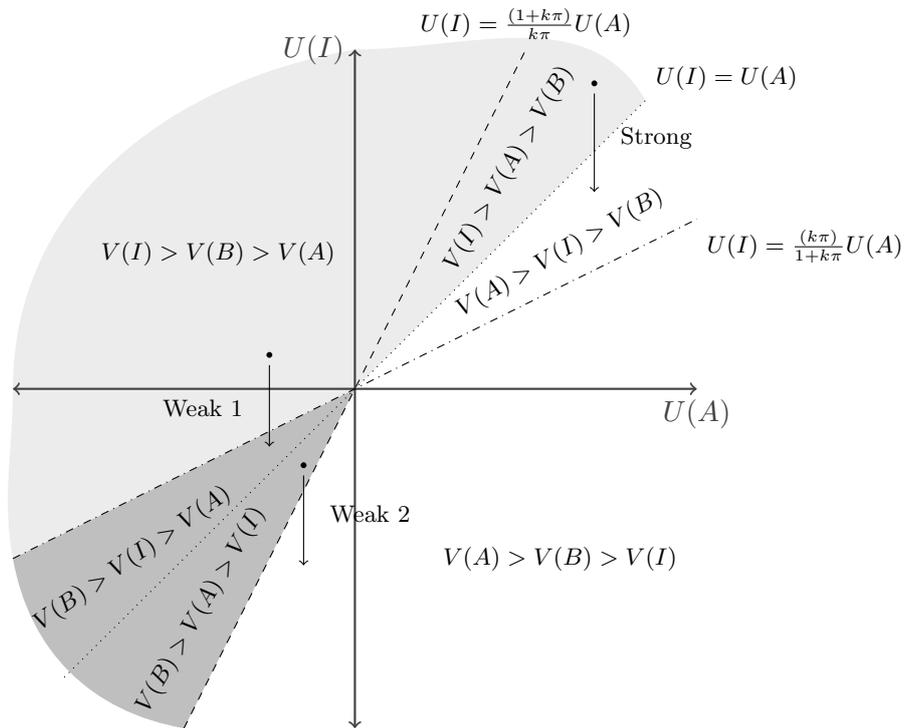
The voter then chooses candidate  $x \in \{I, A, B\}$  who maximizes  $V(x)$ .

Figure 1 clarifies how vote choice depends on voters' "raw" preferences over the candidates (i.e.,  $U(I)$ ,  $U(A)$ , and  $U(B)$ ) when one candidate is uncompetitive and voters are partly expressive and partly instrumental in the manner specified above. In order to characterize a three-way choice in two dimensions, we normalize voter utilities such that  $U(B) = 0$  and characterize voters in terms of  $U(A)$  (horizontal axis) and  $U(I)$  (vertical axis); thus voters whose favorite candidate on expressive grounds is  $B$  are located in the bottom-left quadrant. Now, from Equations 2-4 we know that  $V(I) > V(A)$  when  $U(I) > U(A)$ ;  $V(I) > V(A)$  when  $U(I) > \frac{1+k\pi}{k\pi}U(A)$ ; and  $V(I) > V(B)$  when  $U(I) > \frac{k\pi}{1+k\pi}U(A)$ . These inequalities, indicated by dotted and dashed lines on Figure 1, divide the space into regions according to binary orderings of expected utility. Thus the dark shaded area at the bottom left is where  $V(B) > V(I)$  and  $V(B) > V(A)$ , and therefore these voters will vote for  $B$ ; the white area to the right and above this is where  $V(A) > V(B)$  and  $V(A) > V(I)$  and therefore these voters will vote for  $A$ .

## 2.2 Varieties of punishment for incumbent misconduct

We think of incumbent misconduct as something that reduces  $U(I)$  for voters, which translates into a vertical shift in Figure 1. (This reflects the assumption made above that the raw utilities across candidates are independent.) Figure 1 indicates that with three options there are three ways in which such a vertical shift can affect vote choice: it can move voters from  $I$  to  $A$ ; it

Figure 1: Vote choices as a function of voter preferences



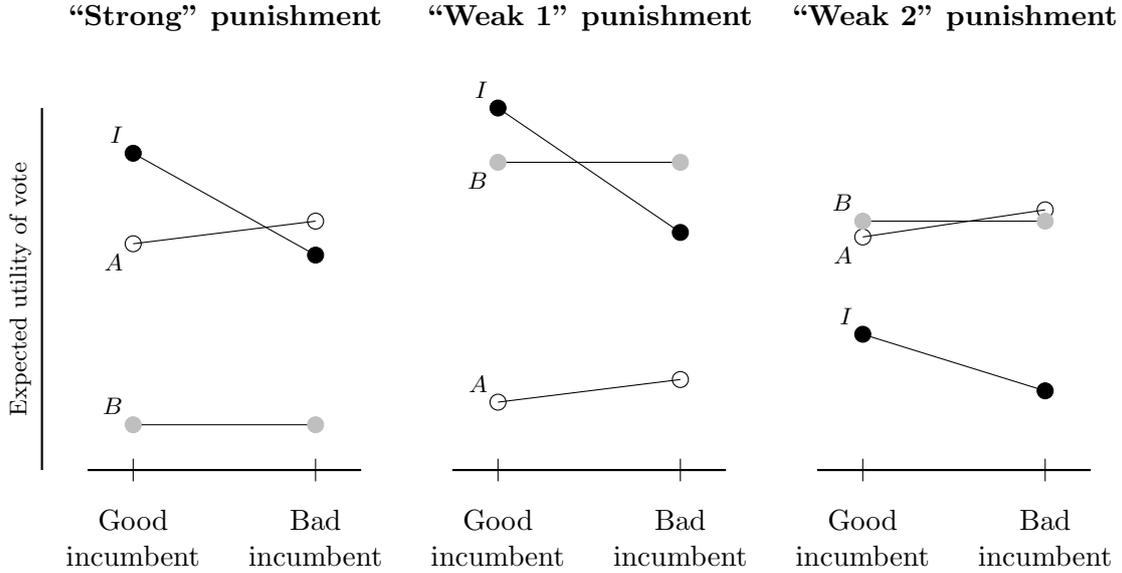
NOTE: By normalizing  $U(B) = 0$ , we can characterize voters with a location in  $U(A)$ - $U(I)$  space. Dotted and dashed lines indicate equal expected utility between voting for two candidates (see text for details). Voters in the light gray area vote for  $I$ , those in the white area vote for  $A$ , and those in the darker gray area vote for  $B$ . The degree to which voters prioritize election outcomes as opposed to the expressive value of voting is given by  $k\pi$ . Three forms of electoral accountability (corresponding to a reduction  $U(I)$  and thus a downward movement) are shown: “Strong” punishment, “Weak 1” punishment, and “Weak 2” punishment.

can move voters from  $I$  to  $B$ ; or it can move voters from  $B$  to  $A$ . We refer to the first shift ( $I \rightarrow A$ ) as “Strong” punishment, the second shift ( $I \rightarrow B$ ) as “Weak 1” punishment, and the third shift ( $B \rightarrow A$ ) as “Weak 2” punishment. (We indicate an example of each on Figure 1.) The distinction between “Strong” and “Weak” punishment becomes clear when we think about the effect of the voter action on election outcomes. We think of the incumbent being primarily concerned with the vote margin between herself and the main challenger; after all, this margin determines (albeit discontinuously) whether she wins re-election. “Strong” punishment reduces this margin by 2 (because the voter shifts from the incumbent to the main challenger) while “Weak” punishment reduces it by 1 (because the voter shifts from the incumbent to an uncompetitive candidate or from an uncompetitive candidate to the main challenger).

Figure 2 illustrates these three forms of punishment in a simpler way. “Strong” punishment comes from the incumbent and challenger changing places at the top of the voter’s ranking of expected utility due to the incumbent’s misconduct, “weak 1” punishment comes from the incumbent plunging below the uncompetitive candidate, and “weak 2” punishment comes from the challenger rising above the uncompetitive candidate. “Strong” punishment is strong because the voter switches from the incumbent to the challenger; “weak” punishment still hurts the incumbent but less so, as the voter switches from the incumbent to an uncompetitive candidate (1) or from an uncompetitive candidate to the challenger (2). These examples correspond to the three cases highlighted in Figure 1.

Figure 2 highlights the role of instrumental voting in electoral accountability when there are multiple candidates on offer. Both “Strong” and “Weak 1” punishment can take place even when voters are purely expressive: the incumbent becomes less attractive and thus is replaced by either the main challenger (“Strong” punishment) or the uncompetitive candidate (“Weak 1” punishment) at the top of the voter’s ordering. “Weak 2”, on the other hand, requires voters to act instrumentally: the incumbent becomes less attractive, such that the voter abandons an uncompetitive candidate for the main challenger. Equation 3 highlights the way in which incumbent misconduct raises the expected utility of voting for the main challenger. This dependence is increasing in  $k\pi$ , which measures the relative importance of instrumental as opposed to expressive motives.

Figure 2: Three forms of punishment



NOTE: Figure shows the voter’s expected utility from voting for the incumbent ( $I$ ), challenger ( $A$ ), or uncompetitive candidate ( $B$ ) as a function of incumbent conduct (“Good” or “Bad”) in three punishment patterns.

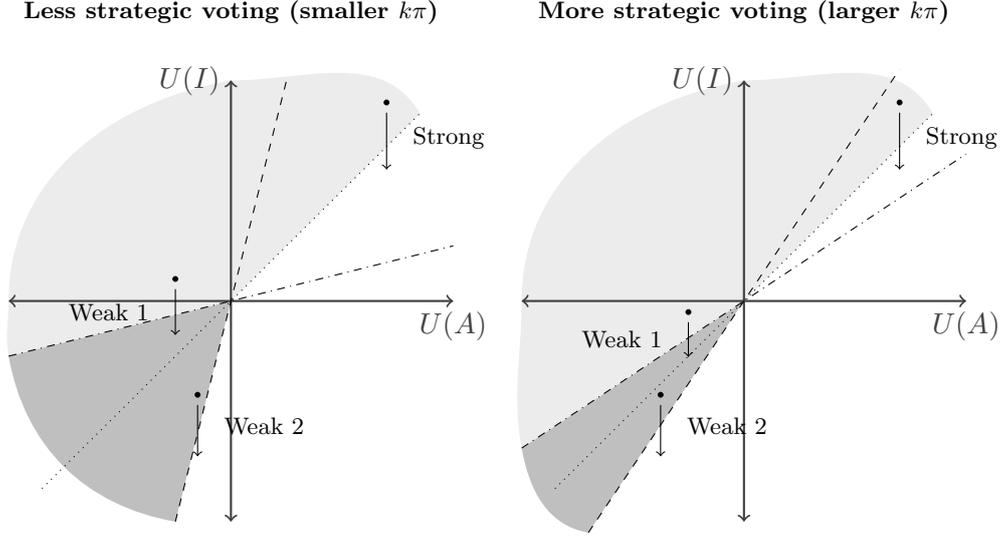
Figure 3 helps to clarify how the possibility for strong and weak punishment relates to  $k\pi$ . In the left panel we assume  $k\pi$  is low, meaning that voting is mostly expressive. Compared to the baseline case in 1, more voters will support  $B$ . More importantly for our purposes, when instrumental voting is low there is little scope for “Weak 2” punishment (assuming a roughly uniform distribution of voters through the space) because of the steepness of the line indicating indifference between  $A$  and  $B$ .<sup>3</sup> By the same logic, the implication is reversed for “Weak 1” punishment, however: the possibility for this form of punishment is greatest when voters are fully expressive. Then again, the scope for “Strong” implication is somewhat smaller when voters are more expressive, as it becomes less likely for a voter to move from the upper-left quadrant of the figure down to a part of the lower-left quadrant in which the voter chooses  $A$ .

### 2.3 Aggregating voter responses to misconduct

In a model assuming that voters choose between an incumbent and one challenger, the effect of some event (such as incumbent misconduct) on electoral outcomes can be summarized by

<sup>3</sup>More formally, assuming voters are distributed roughly uniformly in a circular region about the origin, the region in which voters would engage in “Weak 2” punishment for a given negative shock to  $U(I)$  is roughly a trapezoid whose height, and thus area, is decreasing in  $k\pi$ .

Figure 3: Electoral accountability and strategic voting



NOTE: See note to Figure 3; this figure shows how vote choice and the possibility for strong and weak punishment depends on the extent of strategic voting, as indicated by  $k\pi$ .

Figure 4: Transition matrices indicating the proportion of voters who shift their support from the row option to the column option in response to incumbent misconduct

4.A: Generic transition matrix

$$\begin{matrix} & I & A & B \\ \begin{matrix} I \\ A \\ B \end{matrix} & \begin{pmatrix} \pi_{II} & \pi_{IA} & \pi_{IB} \\ \pi_{AI} & \pi_{AA} & \pi_{AB} \\ \pi_{BI} & \pi_{BA} & \pi_{BB} \end{pmatrix} \end{matrix}$$

4.B: Transition matrix when  $U(I)$  drops

$$\begin{matrix} & I & A & B \\ \begin{matrix} I \\ A \\ B \end{matrix} & \begin{pmatrix} 1 - \alpha - \beta & \alpha & \beta \\ 0 & 1 & 0 \\ 0 & \gamma & 1 - \gamma \end{pmatrix} \end{matrix}$$

just one parameter: the change in the proportion of voters who support the incumbent. More parameters are necessary in a more general setting where voters have multiple options. A natural approach to characterize changes in voter behavior in such settings is to use a transition matrix that characterizes the proportion of voters who move from a given vote choice in one state of the world to another vote choice in another state of the world. Thus in Figure 4.A we provide a generic transition matrix with vote options  $I$ ,  $A$ , and  $B$ ; for example,  $\pi_{IA}$  indicates the probability of supporting  $A$  in the second state of the world, conditional on supporting  $I$  in the first state of the world. (Thus entries in a given row add to one.)

Suppose that the two states of the world are “Good incumbent” and “Bad incumbent”; thus the transition we are considering indicates how voters would respond in aggregate to incumbent misconduct. (For example,  $\pi_{IA}$  would indicate the proportion of voters who, conditional on

supporting the good incumbent, would support the challenger if the incumbent were bad.) Under the assumptions about voter behavior introduced above, we can then simplify the transition matrix to look like 4.B, where  $\alpha$  denotes the proportion of “Strong” punishers among those who would support the good incumbent,  $\beta$  denotes the proportion of “Weak 1” punishers among those who would support the good incumbent, and  $\gamma$  denotes the proportion of “Weak 2” punishers among those who would support the good incumbent. A transition with a zero indicates a behavior that is inconsistent with our assumptions about voter preferences, e.g. voting for  $A$  when the incumbent is good but voting for  $I$  when the incumbent is bad.<sup>4</sup>

Given the proportions in Figure 4, we can characterize the relationship between baseline support for each candidate (meaning support in the absence of incumbent misconduct) and electoral accountability in response to incumbent misconduct. Denote by  $V_x(0)$  the vote share for candidate  $x \in \{I, A, B\}$  in the absence of incumbent misconduct. Then the punishment for incumbent misconduct (in terms of the margin between the incumbent and the challenger) can be expressed as

$$P = (2\alpha + \beta)V_I(0) + \gamma V_B(0). \quad (5)$$

Now, consider how the degree of punishment varies with  $V_B(0)$ . If we denote by  $m$  the margin between the incumbent and main challenger, then we can write  $V_I(0) = \frac{1+m-V_B(0)}{2}$ . Substituting this into Equation 5 and differentiating with respect to  $V_B(0)$  while holding  $m$  fixed,<sup>5</sup> we have that

$$\frac{\partial P}{\partial V_B(0)} = \gamma - \alpha - \beta/2. \quad (6)$$

Equation 6 clarifies how electoral punishment for misconduct depends on baseline support for uncompetitive candidates, holding fixed the margin between the incumbent and the main challenger and assuming that  $\alpha$ ,  $\beta$ , and  $\gamma$  are independent of the baseline vote shares. Thus as competition becomes less Duvergerian, punishments get larger only if  $\gamma > \alpha + \beta/2$ . For example, comparing a situation where, in the absence of misconduct,  $I$  would win 55% of the vote

<sup>4</sup>The transition matrix thus reflects a monotonicity assumption, which is that incumbent misconduct will not make it more attractive to vote for the incumbent or the uncompetitive challenger.

<sup>5</sup>The assumption that  $m$  is held fixed amounts to a requirement that an increase in support for  $B$  comes equally at the expense of  $I$  and  $A$ . It should be clear that in a given setting an increase in support for minor candidates may come at the expense of the incumbent rather than the challenger or vice versa, which implies that it affects the margin between the leading candidates and may imply different conclusions about accountability.

and  $A$  would win 45% of the vote (i.e. two-party competition) to a situation where  $I$  would win 45% of the vote,  $A$  would win 35% of the vote, and  $B$  would win 20% of the vote (multiparty competition), we know that incumbent punishment would be larger in the latter situation only if the proportion of  $B$  supporters who would switch to  $A$  is larger than the proportion of  $I$  supporters who would switch to  $A$  plus half the proportion of  $I$  supporters who would switch to  $B$ .

It is important to be clear about the counterfactual we can learn about from this analysis. Equation 6 does not tell us what would happen to accountability if  $V_B(0)$  rose because of an increase in expressive voting or due to a generic shift in voter preferences toward party  $B$ . Such changes would likely also change  $\alpha$ ,  $\beta$ , and  $\gamma$ , in which case Equation 6 would be incorrect. Neither does it tell us what would happen if support moved from the main challenger to an uncompetitive candidate. Rather, Equation 6 tells us what would happen if there were more of the kinds of voters who support  $B$  in the baseline condition and proportionately fewer of the kinds of voters who support  $I$  and  $A$ . That is, it tells us about a counterfactual in which we inflate the proportion of voters in the dark region of Figure 1 and decrease the proportion of voters in the other two regions of Figure 1, without changing the distribution of preferences within each region; it asks how the effect of a drop in  $U(I)$  in that counterfactual compares to the same effect with the original distribution of preferences.

How can we estimate  $\alpha$ ,  $\beta$ , and  $\gamma$ ? Without imposing additional identifying restrictions we cannot estimate these parameters based on aggregate electoral results, even if we were able to overcome the challenge of observing both  $V_x(0)$  and  $V_x(1)$  for  $x \in \{I, A, B\}$ .<sup>6</sup> The parameters can be estimated with individual data as long as we can produce valid counterfactuals, i.e. vote choices by individuals who are similar but faced different levels of incumbent misconduct. Our approach is to use a survey experiment in which we observe the same respondent's vote choice in both states of the world; this allows us to straightforwardly estimate the parameters  $\alpha$ ,  $\beta$ ,

---

<sup>6</sup>In particular, we have the following three equations describing the vote shares

$$\begin{aligned} V_I(1) &= (1 - \alpha - \beta)V_I(0) \\ V_A(1) &= \alpha V_I(0) + V_A(0) + \gamma V_B(0) \\ V_B(1) &= \beta V_I(0) + (1 - \gamma)V_B(0). \end{aligned}$$

but when we recognize that  $V_B(1) = 1 - V_I(1) - V_A(1)$  and  $V_B(0) = 1 - V_I(0) - V_A(0)$ , it turns out that there are three unknowns and only two independent equations.

and  $\gamma$  for the set of voters in our study, as described in the next section.

### 3 Experimental design

We carried out an online population-based survey experiment (Mutz, 2011) to assess British voters' responses to incumbent misconduct in a variety of scenarios. The experiment was fielded to a sample of 1962 British voters on 2 and 3 June 2014 by YouGov. These respondents were drawn from an online panel of over 360,000 people via YouGov's active sampling algorithm, generating a sample that is representative of the British voting-age population in terms of age, gender, region, social grade and newspaper readership. In terms of both distributions of key demographic variables and parameter estimates in vote choice regressions, there is little evidence of substantial statistical differences between the samples provided by reputable electronic survey companies (including YouGov in the UK) and those from traditional face-to-face probability samples (Sanders et al., 2007).

In our survey experiment, we present respondents with an introductory screen and then a series of five choice tasks. Figure 5 provides screenshots of the introductory screen and an example of one choice task that a respondent could have faced.

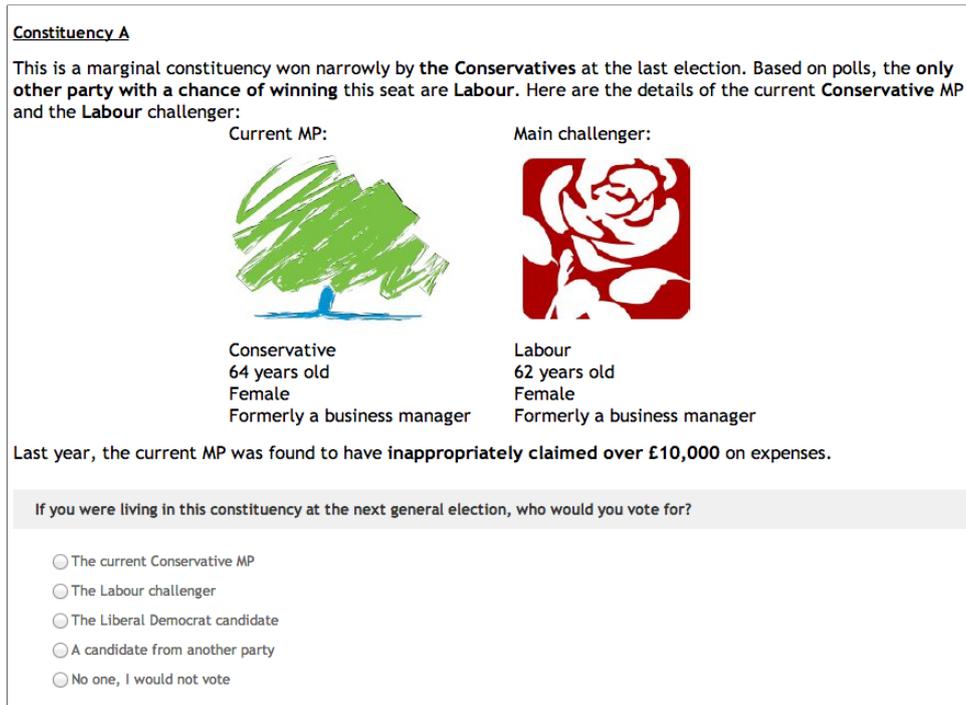
In each choice task, the respondent was presented with a hypothetical constituency contest at the next general election and asked how they would vote if they lived in that constituency. We begin by informing the respondent that the hypothetical constituency is a highly competitive two-way contest between the incumbent MP and an acknowledged 'main challenger', with no other candidates having any realistic chance of winning in the constituency given local opinion polls. We thus restrict our experiment to contexts where it is clear to the respondent that there is a non-trivial probability that their vote could alter the local election outcome (i.e.  $\pi$  is large, in the notation of our theoretical framework) and where it is clear which two candidates have a chance of winning the seat. In contrast with observational studies, this allows us to examine vote choices in an environment where the strategic electoral context as perceived by the voter is likely to be very similar to that measured by the researcher.

We then present the respondent with information about the incumbent MP and the main challenger in the constituency. First, the respondent is given each candidate's *party affilia-*

Figure 5: Screenshots from survey experiment



(a) Introductory screen



(b) Hypothetical election scenario with MP misconduct

NOTE: After the introductory screen, each respondent saw five scenarios. We randomly varied whether the incumbent was guilty of misconduct or not (in the latter cases, the text immediately preceding the survey question read, “Last year, the current MP received a commendation for diligent and ethical service from a Westminster watchdog”) as well as the party matchup (with the four possibilities listed in Table 1) and the age, gender, and former occupation of the candidates.

Table 1: Number of observations for each combination of key experimental treatments

Party match-up (incumbent-challenger)	MP misconduct	
	Good	Bad
Conservative-Labour	1240	1216
Conservative Liberal Democrat	1211	1187
Labour-Conservative	1256	1204
Labour Liberal Democrat	1220	1276

tion (displayed prominently with the help of a party logo). As Table 1 shows, we randomized among four possible party match-ups with equal probability assigned to each one: Conservative incumbent, Labour challenger; Conservative incumbent, Liberal Democrat challenger; Labour incumbent, Conservative challenger; and Labour incumbent, Liberal Democrat challenger. Thus, in a given choice task, a respondent was equally likely to be faced with a contest between a Conservative and a Labour candidate and a contest between a candidate from one of these two parties and a Liberal Democrats candidate.<sup>7</sup>

Next, the respondent is informed about the age, gender, and previous occupation of both the incumbent and the challenger. The values of each of these three attributes were randomized independently.<sup>8</sup>

Finally, the respondent was given information about the incumbent MP’s ethical conduct. This MP attribute, which we refer to as the “MP misconduct treatment”, took on two possible values. In the “good MP” treatment condition, the respondent was informed that “Last year, the current MP received a commendation for diligent and ethical service from a Westminster watchdog”. In the “bad MP” treatment condition, the respondent was informed that “Last year, the current MP was found to have inappropriately claimed over £10,000 in expenses.” Assignment to good or bad MP treatment in a given choice task was randomized independent of all other MP or challenger attributes, and of all attributes in previous choice contexts that the respondent faced.

At the bottom of the choice task screen, the respondent is given a set of vote choices. Importantly, we give respondents the option of voting for the current MP, the challenger, another

<sup>7</sup> Note that the incumbent is always either Conservative or Labour, while the challenger can be Conservative, Labour, or Liberal Democrat.

<sup>8</sup> The possible ages were 45, 52 and 64 for the incumbent and 40, 52 and 64 for the challenger. The possible previous occupations were: a GP, a journalist, a political advisor, a teacher and a business manager.

party’s candidate or abstention.

Because we simultaneously randomize several attributes of the choices available to respondents, our design is similar to that of a conjoint analyses. Conjoint analysis is increasingly being applied in political science, and our design shares some of the advantages of this approach (Hainmueller, Hopkins and Yamamoto, 2014). The design presents respondents with a multidimensional and thus reasonably realistic choice setting, going beyond a simple misconduct cue. This should increase external validity and also reduce social desirability bias, since respondents can justify their vote based on a number of considerations rather than just the conduct of the incumbent (Hainmueller, Hopkins and Yamamoto, 2014).

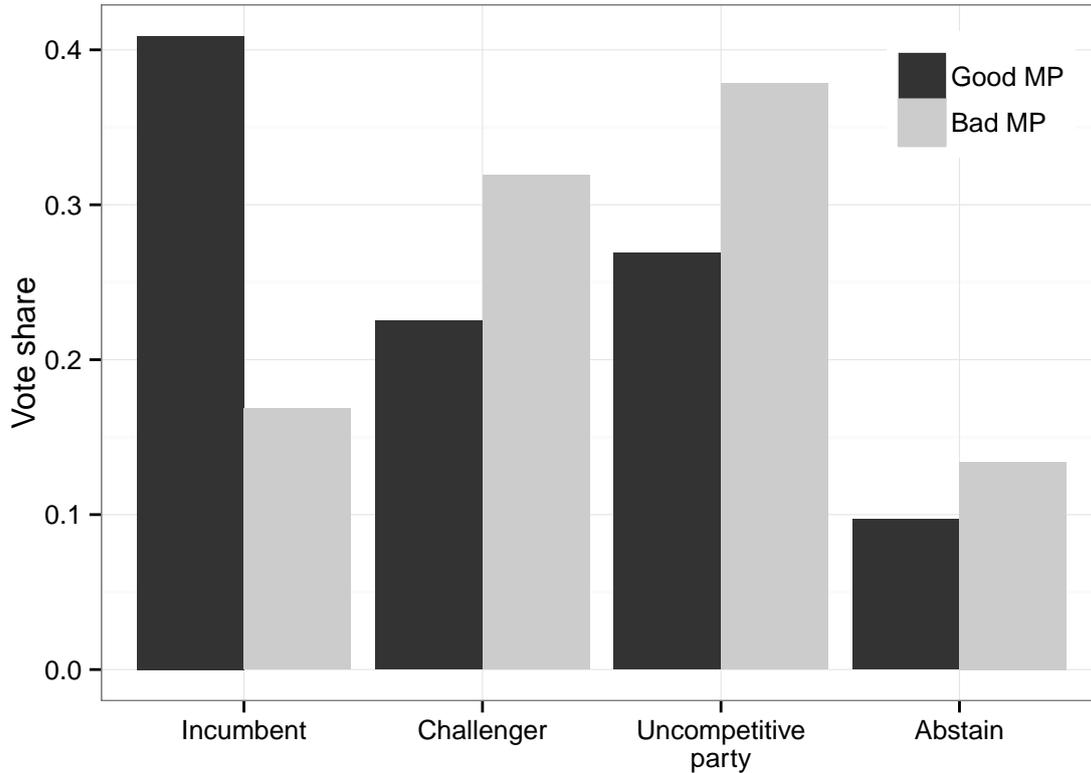
We also try to enhance the external validity of the findings by priming respondents to think about partisan general election considerations. We do so in the introductory screen by emphasising the national-government-selection role of a general election and linking this to policy differences between the three nationally established parties – and party leaders – competing for a role in government. We also do this in the individual choice tasks by prominently displaying the party logo of the incumbent and main challenger. Of course, partisan considerations will naturally be weaker in an experimental setting than in a real-world general election campaign where the stakes are higher. But we feel that our design goes some way to attenuating this problem and therefore allowing us to induce voters to trade off partisanship and concerns about misconduct in a realistic manner.

More generally, our experimental approach ensures that we have detailed control over the information respondents receive about the conduct of the MP and about the strategic electoral context. In contrast, in an observational setting we have to make a stronger assumption that voter perceptions of incumbent misconduct and strategic context reflect the objective measures we have access to as researchers.

## 4 Results

Figure 6 compares the distribution of vote choices across MP misconduct treatments, averaging over all observations and thus all possible MP and challenger attributes in the experiment. It is immediately clear from the plot that support for the incumbent drops substantially when

Figure 6: Vote choice by incumbent MP misconduct treatment



NOTE: Distribution of observed vote choices by MP misconduct treatment, averaging over all observations (i.e. all respondents and all possible MP and challenger attributes in the experiment). In the ‘good MP’ treatment, respondents were told that, “Last year, the current MP received a commendation for diligent and ethical service from a Westminster watchdog”. In the ‘bad MP’ treatment, respondents were told that, “Last year, the current MP was found to have inappropriately claimed over £10,000 in expenses.” A vote for a candidate from any party other than that of the incumbent MP or the main challenger is coded as an “uncompetitive party” vote.

he or she over-claims on expenses (the ‘bad MP’ treatment), compared to when he or she is commended for diligent and ethical service (the ‘good MP’). Voters opted for the incumbent MP 41 per cent of the time in the commendation treatment, but only 17 per cent of the time in the misconduct treatment. This reduction in support of 24 per cent is strongly significant ( $t = -22.7$ ), and is much larger than the average effect of MP implication in the expenses scandal (Pattie and Johnston, 2012). This is unsurprising, since in our experiment voters are fully informed about the behaviour of their MP, whereas at the 2010 election there is evidence that voter perceptions of local MP misconduct were subject to considerable error (Vivyan, Wagner and Tarlov, 2012).

Consistent with our discussion of the different types of punishment for misconduct, Figure 6 also makes clear that misconduct does not simply lead to a transfer of support from incumbent to the competitive challenger. Instead, misconduct causes a significant increases in support for uncompetitive parties (11 per cent,  $t = 10.5$ ) and abstention (3.7 per cent,  $t = 5.3$ ), as well as increases in support for the competitive challenger (9.4 per cent,  $t = 9.6$ ). It is not immediately clear from this Figure alone how misconduct drives the changes in the vote choice distribution. The rise in votes for the main challenger may be driven by “Strong” or “Weak 2” misconduct punishment, whereas the change in votes for uncompetitive parties reflects the combination of “Weak 1” punishment (which increases support for uncompetitive parties) and “Weak 2” punishment (which decreases it).

In what follows, we combine support for uncompetitive candidates with abstention; thus in terms of the framework we have support for the incumbent ( $I$ ), support for the competitive challenger ( $A$ ), and everything else ( $B$ ). The results are essentially the same if we omit abstainers or treat them as a separate category, though in the latter case the analysis becomes somewhat more complicated.

#### 4.1 Estimating the proportion of strong and weak punishers

We now exploit the fact that our experiment yields repeated observations for each respondent in order to estimate the proportion of “Strong”, “Weak 1” and “Weak 2” punishers in each party match-up for the set of voters in our study. Specifically, for each possible party match-up in the experiment - Conservative incumbents facing Labour and Liberal Democrat challengers and Labour incumbents facing Conservative and Liberal Democrat challenger - we do the following.

First, we subset our data to those respondents who are confronted with a given seat-type at least once with a “good MP” and at least once with a “bad MP”. These are the respondents for whom we have valid counterfactuals.<sup>9</sup> This leaves us with sub-samples of between 394 and 439 across the four party match-ups. Because our party match-up and MP misconduct treatments are randomised, these sub-samples are random samples from the full set of voters.

---

<sup>9</sup>We could increase our sample size and power by opting for a less strict criteria for defining counterfactual observations. For example, we could use match observations based on respondent background characteristics. We plan to explore this in future drafts of the paper.

Figure 7: Estimated generic transition matrices by party contest

7.A: Conservative-Labour				7.B: Conservative-Liberal Democrat			
	I	A	B		I	A	B
I	0.413	0.148	0.439	I	0.385	0.252	0.363
A	0.026	0.903	0.071	A	0.019	0.815	0.167
B	0.010	0.152	0.838	B	0.009	0.098	0.893

7.C: Labour-Conservative				7.D: Labour-Liberal Democrat			
	I	A	B		I	A	B
I	0.434	0.114	0.451	I	0.361	0.333	0.306
A	0.018	0.936	0.045	A	0.020	0.860	0.120
B	0.031	0.085	0.884	B	0.012	0.071	0.918

Second, we use this sub-sample of respondents for whom we observe a counterfactual to estimate a generic transition matrix like that defined in Figure 4, which quantifies the proportion of voters who shift support from a given vote choice to another vote choice in response to MP misconduct. This transition matrix is generated using the fitted probabilities from a simple multinomial logistic regression where the response variable is  $V(1) \in \{I, A, B\}$  and the only predictor is  $V(0) \in \{I, A, B\}$ .<sup>10</sup>

The estimated transition matrices appear in Figure 7. The first row in each matrix indicates how respondents who voted for the “good” incumbent behaved when faced with a “bad” incumbent of the same party, facing a challenger of the same party. For example, in cases with a Conservative incumbent and Labour challenger, about 40% of respondents who voted for the good incumbent also voted for the bad incumbent, about 15% switched to the challenger (which is “Strong” punishment), and about 45% switched to an uncompetitive candidate or chose not to vote (which is “Weak 1” punishment). Among incumbent supporters, the proportion who engage in “Strong” punishment ranges across matchups from 0.114 (Labour-Conservative matchups) to 0.333 (Labour-Liberal Democrat matchups); consistent with the logic and evidence of Eggers (Forthcoming), the rate of “Strong” punishment is higher in the contests involving a Liberal

<sup>10</sup> For the Conservative-Liberal Democrat sub-sample we have an empty cell problem: there are no respondents who vote for  $A$  when the MP is good, and  $I$  when the incumbent is bad. To ensure stable multinomial regression estimates for this sub-sample, we add 9 extra ‘pseudo-observations’, one for each possible combination of  $V(0)$  and  $V(1)$  values.

Democrat challenger, suggesting that Labour and Conservative supporters tend to see the Liberal Democrats as a relatively palatable, “middle ground” alternative. In all four contexts, we also see a considerable amount of “Weak 1” punishment; indeed, the proportion engaging in “Weak 1” punishment is larger than the proportion engaging in “Strong” punishment (i.e.  $\beta > \alpha$ ) in three of the four scenarios, and substantially larger in the Conservative-Labour and Labour-Conservative contests. This highlights an important part of electoral accountability that would be missed in a binary approach that focuses only on support for the incumbent: comparing the Conservative-Labour contest and the Conservative-Liberal Democrat contest, for example, a similar proportion of respondents defect from the incumbent in the two cases, but the punishment suffered by the incumbent is more severe in the Conservative-Liberal Democrat case because substantially more of the defectors move to the challenger rather than support an uncompetitive candidate or abstain.

We now turn to the second row of the transition matrices. Here our expectation was that the second element of this row would be 1, meaning that all voters who support the challenger when the incumbent is good should do so when the incumbent is bad. (In technical terms, we expect voting for  $A$  to be an “absorbing state”.) In fact, we see a few respondents moving to the incumbent and a substantial proportion (ranging from about .05 to .17) moving from the challenger to an uncompetitive candidate or abstention. Indeed, in Conservative-Liberal Democrat contests and Labour-Liberal Democrat contests the proportion of challenger supporters who move to an uncompetitive candidate is larger than the proportion of uncompetitive supporters who move to the challenger (i.e. “Weak 2” punishers). This pattern suggests that in some cases incumbent misconduct taints the challenger, such that the voter prefers to abstain or cast a protest vote than to vote for the only candidate with a chance of removing the bad incumbent (even though the same voter preferred the challenger to the good incumbent). This seems to happen especially when the challenger is a Liberal Democrat, which suggests that the perceived proximity of the Liberal Democrats to both major parties is a double-edged sword when it comes to electoral accountability: on the one hand, more voters are willing to defect from a Labour or Conservative incumbent to the challenger when that challenger is a Liberal Democrat; on the other hand, some voters who supported a Liberal Democrat challenger see the

Liberal Democrats as somehow implicated in incumbent misconduct and respond by abstaining or casting a protest vote for another party. Put simply, the Liberal Democrat challenger may be tainted by incumbent corruption precisely because voters view the Liberal Democrat challenger as closer to the incumbent.<sup>11</sup>

Turning to the third row of the transition matrices, we see that only between 7% and 15% of respondents who support an uncompetitive party or abstain in the baseline case switch to the main challenger in the event of incumbent misconduct. Put differently, across contexts we see that at least 85% of the respondents who supported an uncompetitive candidate or abstained when faced with a good incumbent continued to do so when faced with a bad incumbent. This is a first indication that increasing the baseline support for uncompetitive candidates is unlikely to increase punishment for incumbent misconduct. In the next section we use these estimates to look more systematically at the relationship between support for uncompetitive candidates and electoral accountability.

## 4.2 Measuring the effect of non-Duvergerian competition

We now consider a counterfactual experiment in which we vary the proportion of voters supporting the uncompetitive candidate (or abstaining) while holding fixed the margin between the incumbent and challenger; this is equivalent to raising the number of voters supporting the uncompetitive candidate while proportionately reducing the number of voters supporting the incumbent and challenger. Equation 6 indicated the relationship between baseline support for candidate  $B$  and the level of punishment in our model under the monotonicity assumption that incumbent misconduct does not attract anyone to the incumbent or away from the challenger; in terms of the general transition matrix in Figure 4, that equation states that

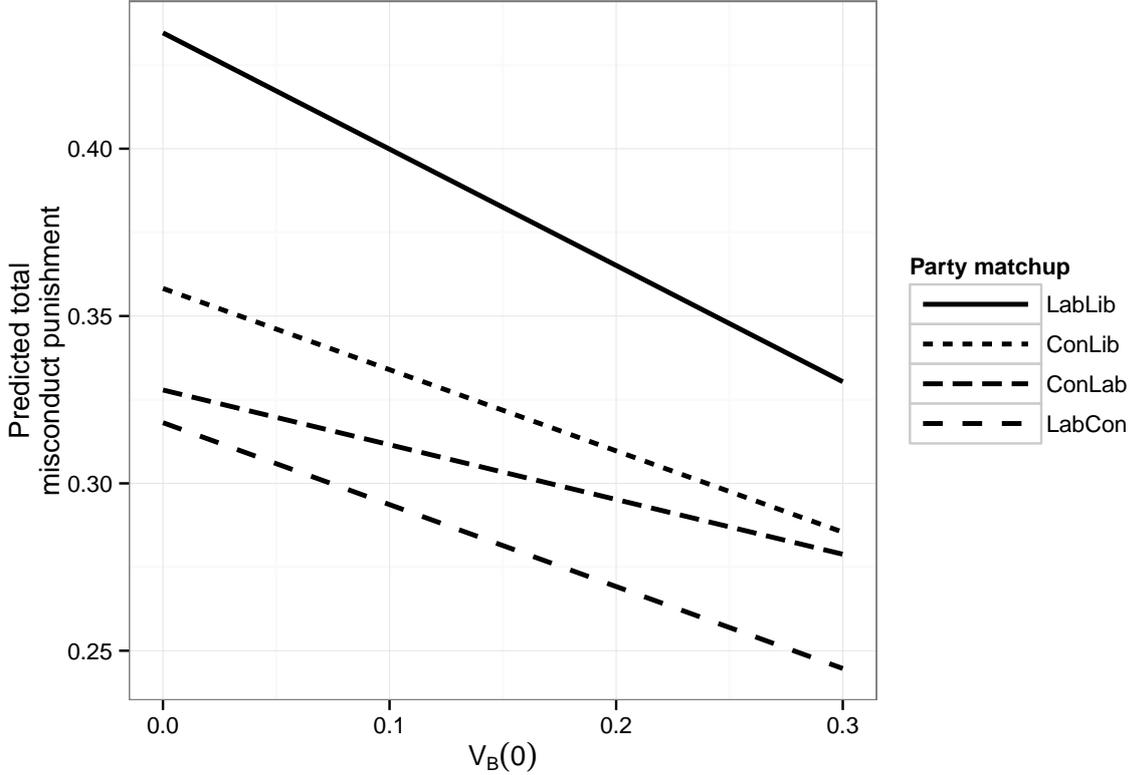
$$\frac{\partial P}{\partial V_B(0)} = \pi_{BA} - \pi_{IA} - \pi_{IB}/2.$$

As noted above, in the experiment a small number of respondents did in fact engage in other forms of behavior, particularly moving from  $A$  to  $B$ . Rather than ignore that behavior, we take

---

<sup>11</sup>This may occur in part because respondents partly blame the Liberal Democrats, who were in a governing coalition with the Conservatives at the time of the survey, for regulating MP behavior in the House of Commons.

Figure 8: Punishment as a function of seat type and  $V_B(0)$



NOTE: Predicted total punishment as a function of party match-up and baseline uncompetitive voting ( $V_B(0)$ ). Predicted total punishment is calculated by taking the estimated  $V(0) \rightarrow V(1)$  transition matrix for each party match-up and post-stratifying to multiple artificial populations whose proportion of baseline uncompetitive voters varies. The baseline incumbent-challenger vote margin ( $V_I(0) - V_A(0)$ ) is held constant at 0.05.

it into account and estimate:

$$\frac{\partial P}{\partial V_B(0)} = \pi_{BA} - \pi_{IA} - \pi_{IB}/2 + \pi_{AI} + \pi_{AB}/2 - \pi_{BI}. \quad (7)$$

Of the extra elements, only  $\pi_{AB}$  was noticeably above zero in our experiment; to the extent that some baseline supporters of  $A$  move to  $B$  in response to misconduct (which tends to *reward* incumbent misconduct), then an increase in baseline supporters of  $B$  will reduce the proportion of voters who engage in this behavior, which (all else equal) increases the overall punishment.

Figure 8 shows the relationship between baseline support for  $B$  and overall punishment implied by Equation 7 for each match-up. (The intercept is given by  $2 \times \pi_{IA} + \pi_{IB} - 2 \times \pi_{AI} - \pi_{AB}$ .) The figure indicates that for all four seat types the degree of punishment is

decreasing in the baseline support for uncompetitive parties, holding fixed the margin between the incumbent and challenger. An increase of 10 percentage points in support for uncompetitive candidates reduces the punishment by between 2 and 3 percentage points, depending on the seat type. For all four party match-ups, the 95% confidence interval on the slope (which we derive through simulation<sup>12</sup>) does not include zero. (For Conservative-Labour matchups the interval is  $[-0.27, -0.06]$ ; for Conservative-Liberal Democrat matchups it is  $[-0.34, -0.11]$ ; for Labour-Conservative matchups it is  $[-0.32, -0.14]$ ; and for Labour-Liberal matchups it is  $[-0.42, -0.21]$ .)

The figure also indicates that the level of punishment is higher in Conservative-Liberal Democrat matchups and (especially) Labour-Liberal Democrat matchups. Inspection of the transition matrices in Figure 7 indicates that this is mainly because  $\alpha$  is higher in constituencies with a Liberal Democrat challenger, meaning that a higher proportion of incumbent supporters switch to the main challenger. This is what one would expect if, as indicated by Eggers (Forthcoming), voters tend to view the Liberal Democrats as a kind of “middle ground” between Labour and the Conservatives.

## 5 Discussion and conclusion

This paper has examined how electoral punishment for misconduct operates in a multiparty contest. We provided a theoretical framework showing that punishment can take three forms in such contests: “Strong” punishment, where a voter votes for the incumbent when the incumbent is good, but votes instead for the main challenger when the incumbent is bad; “Weak 1” punishment, where a voter votes for the incumbent when the incumbent is good, but instead votes for an uncompetitive candidate when the incumbent is bad; and “Weak 2” punishment, where a voter votes for an uncompetitive candidate when the incumbent is good, but votes instead for the main challenger when the incumbent is bad. Our approach contrasts with stan-

---

<sup>12</sup>Specifically, we take the multinomial regression model used to generate a given transition matrix in Figure 7 and simulate 1000 draws from the joint posterior distribution of the model coefficients, which we approximate using a multivariate normal distribution with means equal to the coefficient point estimates and covariance matrix equal to the estimated regression covariance matrix. For each draw of the coefficients, we calculate a simulated transition matrix and, based on this, an estimated value of  $\frac{\partial P}{\partial V_B(0)}$ ; the confidence intervals indicate the .025 and .975 quantiles of this distribution of estimates.

standard two-party models of electoral accountability, where punishment can occur solely through “Strong” punishment. We then wrote down a condition indicating whether more baseline support for uncompetitive parties increases or decreases electoral accountability (holding fixed the margin between the incumbent and the challenger) as a function of the proportion of incumbent supporters and uncompetitive-candidate supporters engaging in each type of punishment. Our experiment allowed us to estimate these parameters for a representative sample of British voters in a given hypothetical scandal. The results indicate that across common party match-ups the proportion of uncompetitive candidate supporters who switch to the challenger in response to incumbent misconduct is quite low compared to the proportion of incumbent supporters who switch to the challenger or to an uncompetitive candidate; as a result the estimated effect of a counterfactual increase in baseline support for uncompetitive candidates is to reduce overall punishment for incumbent misconduct. This suggests that, for British voters at least, non-Duvergerian competition tends to undermine the effectiveness of electoral accountability.

While we think the framework and methods we present may be useful in studying electoral accountability in many electoral contexts, we recognize that one should be cautious before applying the conclusions from a survey experiment to actual elections in the UK or elsewhere. We highlight two main challenges to external validity. First, respondents in the experiment we study were faced with an unusually stark episode of incumbent misconduct in the context of a hypothetical election; the magnitude of their response to incumbent misconduct was, not surprisingly, very large (and indeed is much larger than was seen following the expenses scandal on which our misconduct scenarios were based). We expect that the proportion of voters engaging in each form of punishment (“Strong”, “Weak 1”, and “Weak 2”) would be substantially smaller in a more realistic situation in which voters are imperfectly informed about incumbent misconduct and/or have more information about other aspects of the candidates and parties. What is important for our main claim about accountability in non-Duvergerian elections is not the *absolute* magnitude of these proportions but their *relative* magnitude. That is, our claim that higher baseline support for uncompetitive candidates reduces accountability will be valid even for less stark episodes of misconduct as long as the proportion of voters engaging in each form of punishment is proportionally smaller. Future work might assess this by seeing how

these proportions vary with the severity or salience of incumbent misconduct.

The second main challenge to external validity is that our conclusions are necessarily specific to the party system and political context in which we conducted our experiment. The proportion of voters undertaking each form of punishment is not a fundamental and universal feature of human nature, but rather reflects voters' attitudes toward the specific parties fielding candidates, the salience of party as opposed to individual performance in a particular system, and the identity of the parties that are competitive in a given electoral context (as shown by the variation across match-ups that we study). Thus it could be that in another setting the relationship between support for uncompetitive candidates and electoral accountability is neutral or even positive. In particular, in a setting where voters are relatively ambivalent between the challenger and one or more uncompetitive candidates but have strong feelings between the incumbent and challenger it may well be that "Weak 2" punishment would outweigh "Strong" and "Weak 1" punishment, such that an increase in baseline support for uncompetitive candidates would increase overall accountability.

## References

- Cox, G.W. 1997. *Making votes count*. Cambridge University Press Cambridge.
- Eggers, A. Forthcoming. "Partisanship and Electoral Accountability: Evidence from the UK Expenses Scandal." *Quarterly Journal of Political Science* .
- Ferree, Karen E., G. Bingham Powell and Ethan Scheiner. 2014. "Context, Electoral Rules, and Party Systems." *Annual Review of Political Science* 17(1):421–439.  
**URL:** <http://dx.doi.org/10.1146/annurev-polisci-102512-195419>
- Hainmueller, Jens, Daniel J. Hopkins and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(1):1–30.  
**URL:** <http://pan.oxfordjournals.org/content/22/1/1.abstract>
- Mutz, D.C. 2011. *Population-based survey experiments*. Princeton University Press.
- Pattie, C. and R. Johnston. 2012. "The electoral lipact of the UK 2009 MP's expenses scandal." *Political Studies* 60:730–750.
- Raymond, Christopher D. 2013. "Increasingly Irrational Party Systems: Evidence that There Are Increasingly More Than Two Parties at the District Level." *Government and Opposition* 48:506–525.
- Sanders, David, Harold D. Clarke, Marianne C. Stewart and Paul Whiteley. 2007. "Does Mode Matter For Modeling Political Choice? Evidence From the 2005 British Election Study." *Political Analysis* 15(3):257–285.
- Singer, Matthew M. 2013. "Was Duverger Correct? Single-Member District Election Outcomes in Fifty-three Countries." *British Journal of Political Science* 43:201–220.
- Vivyan, N., M. Wagner and J. Tarlov. 2012. "Representative misconduct: voter perceptions and accountability: evidence from the 2009 House of Commons expenses scandal." *Electoral Studies* .